PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C07H 21/04, C07K 14/00, C12N 15/85

A1

(11) International Publication Number: WO 00/32618

(43) International Publication Date: 8 June 2000 (08.06.00)

(21) International Application Number: PCT/US99/28315

(22) International Filing Date: 30 November 1999 (30.11.99)

(30) Priority Data:
09/201,227 30 November 1998 (30.11.98) US

(71) Applicant: EXELIXIS PHARMACEUTICALS, INC. [US/US]; 260 South Littlefield Avenue, South San Francisco, CA 94080 (US).

(72) Inventors: KEYES, Linda, N.; 46 Madera Avenue, San Carlos, CA 94070 (US). DOBERSTEIN, Stephen, K.; 2 Townsend Street, San Francisco, CA 94107 (US). BUCHMAN, Andrew, R.; 3119 Eton Avenue, Berkeley, CA 94705 (US). REDDY, Bindu, P.; 885 Greenwich Street, San Francisco, CA 94133 (US). RUDDY, David, A.; 885 Greenwich Street, San Francisco, CA 94133 (US).

(74) Agents: ANTLER, Adriane, M. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: NUCLEIC ACIDS AND POLYPEPTIDES OF *D. MELANOGASTER* INSULIN-LIKE GENES AND USES THEREOF

(57) Abstract

The present invention relates to *D. melanogaster* insulin–like genes and methods for identifying insulin–like genes. The methods provide nucleotide sequences of *D. melanogaster* insulin–like genes, amino acid sequences of their encoded proteins, and derivatives (*e.g.*, fragments) and analogs thereof. The invention further relates to fragments (and derivatives and analogs thereof) of insulin–like proteins which comprise one or more domains of an insulin–like protein. Antibodies to an insulin–like protein, and derivatives and analogs thereof, are provided. Methods of production of an insulin–like protein (*e.g.*, by recombinant means), and derivatives and analogs thereof, are provided. Further, methods to identify the biological function of a *D. melanogaster* insulin–like gene are provided, including various methods for the functional modification (*e.g.*, overexpression, underexpression, mutation, knock–out) of one or more genes simultaneously. Still further, methods to identify a *D. melanogaster* gene which modifies the function of, and/or functions in a signaling pathway with an insulin–like gene are provided. The invention further provides uses of *Drosophila* insulin–like nucleic acids and proteins, *e.g.*, as media additives, and as pesticides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

NUCLEIC ACIDS AND POLYPEPTIDES OF *D. MELANOGASTER* INSULIN-LIKE GENES AND USES THEREOF

PRIORITY APPLICATION

This application claims priority to U.S. Ser. No. 09/201,227 (Keyes *et al.*) filed November 30, 1998.

BACKGROUND OF THE INVENTION

Insulin is the central hormone governing metabolism in vertebrates (reviewed in Steiner et al., 1989, In *Endocrinology*, DeGroot, eds. Philadelphia, Saunders: 1263-1289). In humans, insulin is secreted by the beta cells of the pancreas in response to elevated blood glucose levels which normally occur following a meal. The immediate effect of insulin secretion is to induce the uptake of glucose by muscle, adipose tissue, and the liver. A longer term effect of insulin is to increase the activity of enzymes that synthesize glycogen in the liver and triglycerides in adipose tissue. Insulin can exert other actions beyond these "classic" metabolic activities, including increasing potassium transport in muscle, promoting cellular differentiation of adipocytes, increasing renal retention of sodium, and promoting production of androgens by the ovary. Defects in the secretion and/or response to insulin are responsible for the disease diabetes mellitus, which is of enormous economic significance. Within the United States, diabetes mellitus is the fourth most common reason for physician visits by patients; it is the leading cause of end-stage renal disease, non-traumatic limb amputations, and blindness in individuals of working age (Warram et al., 1995, In *Joslin's Diabetes Mellitus*, Kahn and Weir, eds., Philadelphia, Lea & Febiger, pp. 201-215; Kahn et al., 1996, Annu. Rev. Med. 47:509-531; Kahn, 1998, Cell 92:593-596). Beyond its role in diabetes mellitus, the phenomenon of insulin resistance has been linked to other pathogenic disorders including obesity, ovarian hyperandrogenism, and hypertension.

Insulin-like proteins are a large and widely-distributed group of structurally-related peptide hormones that have pivotal roles in controlling animal growth, development, reproduction, and metabolism. Consequently, the insulin superfamily has become one of the most intensively investigated classes of peptide hormones. Studies of insulin-like

molecules in invertebrates have been motivated by the desire to identify proteins that play analogous roles to the well-characterized activities of insulin and IGF in mammals. Although insulin superfamily members in invertebrates have been less extensively analyzed than in vertebrates, a number of different subgroups have been defined. Such subgroups include molluscan insulin-related peptides (MIP-I to MIP-VII) (Smit et al., 1988, Nature 331:535-538; Smit et al., 1995, Neuroscience 70:589-596), the bombyxins of lepidoptera (originally referred to as prothoracicotropic hormone or PTTH) (Kondo et al., 1996, J Mol. Biol. 259:926-937), and the locust insulin-related peptide (LIRP) (Lagueux et al., 1990, Eur. J. Biochem. 187:249-254). Most recently, there have been descriptions of an exceptionally large insulin-like gene family in the nematode *C. elegans* (WO1999US08522; Duret, et al., 1998, Genome Res. 8:348-353; Brousseau, et al., 1998, Early 1998 East Coast Worm Meeting, abstract 20; Kawano, et al., 1998, Worm Breeder's Gazette 15(2):47; Pierce and Ruvkun, 1998, Early 1998 East Coast Worm Meeting, abstract 150; Wisotzkey and Liu, 1998, Early 1998 East Coast Worm Meeting, abstract 206). Also, putative orthologs of both vertebrate insulin and IGF have been identified in a tunicate (McRory and Sherwood, 1997, DNA and Cell Biology 116:939-949). From the extensive sequence divergence evident among known subfamilies of insulin-like proteins, it is assumed that this is an ancient family of regulatory hormones that evolved to control growth, reproduction and metabolism in early metazoans. However, the precise evolutionary origins of this important family remain unclear.

Early attempts to propagate *Drosophila* cells in culture revealed a growth factor requirement in defined medium which could be provided by purified bovine insulin, implying the existence of a related endogenous factor in. Also, bovine and human insulin were found to stimulate the differentiation of *Drosophila* cells grown in culture (Seecof and Dewhurst, 1974, Cell Differ. 3(1):63-70; Pimentel, et al., 1996, Biochem. Biophys. Res. Commun. 226(3):855-61). One report described the presence of an "insulin-like activity" in unpurified *Drosophila* extracts that elicited a hypoglycemic effect when injected into mice, although the activity was not particularly strong (Meneses and De Los Angeles Ortiz, 1975, Comp. Biochem. Physiol. A. 51(2):483-5). Another group (LeRoith, et al., 1981, Diabetes 30(1):70-6) fractionated an insulin-like material from *Drosophila* based on immunoreactivity and showed that this material had insulin-like activity on isolated rat adipocytes. Also, polyclonal antibodies raised against bovine/porcine insulin were used to localize insulin-immunoreactive material in *Drosophila* tissue (Gorczyca, et

al., 1993, J. Neurosci. 13(9):3692-704), and specific insulin-immunoreactive substances were detected at neuromuscular junctions and in the central nervous system. However, these substances were not characterized further to validate that they correspond to *bona fide* insulin proteins at the level of primary protein sequence. Indeed, despite this long history of phenomenological evidence for insulin-like activities, true insulin-like genes and proteins in *Drosophila* have not been identified and characterized at the sequence level.

More compelling evidence for evolutionary conservation of insulin-like signaling pathways in *Drosophila* has come from the identification of an apparent homolog of the insulin receptor (Petruzzelli et al., 1986, Proc. Natl. Acad. Sci. U.S.A. 83:4710-4714). One insulin receptor homolog has been characterized thus far in *Drosophila*, termed InR (insulin receptor) also known as DIR (*Drosophila* insulin receptor) (Ruan et al., 1995, J. Biol. Chem. 270:4236-4243), which exhibits extensive homology with vertebrate insulin and IGF receptors in both the extracellular ligand-binding domain and the intracellular tyrosine kinase domain. Genetic analysis of InR function in *Drosophila* has revealed that it is an essential gene with an apparent role in the development of the epidermis and nervous system, as well as growth control (Fernandez et al., 1995, EMBO J. 14:3373-3384). Flies that are homozygous for mutations in InR generally exhibit an embryonic lethal phenotype, but flies bearing certain heteroallelic combinations of InR mutations live to adulthood and the surviving animals have about 50% the normal body weight (Garafalo, Chen, et al., 1996, Endocrinology 137(3):846-56). This result is reminiscent of a similar phenotype observed in mutant mice lacking functional IGF-I receptor genes (Liu, et al., 1993, Cell 75(1):59-72). Aside from this potential role of InR in growth regulation, the role, if any, that InR may have in metabolic regulation in *Drosophila* remains unclear. The ligand binding specificity of InR has been examined using in vitro assays for receptor activation/phosphorylation, and competitive binding of test ligands compared to porcine insulin (Fernandez-Almonacid and Rosen, 1987, Mol. Cell Biol. 7(8):2718-27). Curiously, the results of this study indicated that InR binds vertebrate insulin, and does not apparently recognize vertebrate IGF-I or IGF-II, or even bombyxin-II from the silkworm, implying that the natural *Drosophila* ligands for InR may bear more structural resemblance to vertebrate insulin than to these other insulin superfamily proteins.

The structural homologies of components of the Drosophila *InR* pathway with those involved in insulin signaling in mammals, as well as the function of the *InR* pathway

in controlling growth, and the circumstantial evidence for *Drosophila* insulin-like activities, raise critical questions with respect to further analysis of this pathway and its potential applications. Important issues regarding the biological function, regulation, and signaling mechanisms of insulin superfamily hormones could best be addressed if these pathways could be analyzed using model genetic organisms. In particular, the facile genetic tools currently available in two model organisms, the fruit fly *Drosophila melanogaster* and the nematode *Caenorhabditis elegans*, have proven to be of enormous utility in defining the biological function of genes through mutational analysis, as well as for identifying the components of biochemical pathways conserved during evolution with large-scale, systematic genetic screens (Scangos, 1997, Nature Biotechnol. 15:1220-1221; Miklos and Rubin, 1996, Cell 86:521-529). Key discoveries regarding constituents of a number of important human disease pathways, such as the Ras pathway and the pathway controlling programmed cell death, first came from genetic analysis of pathways known to have an evolutionary relation in *Drosophila* and *C. elegans*, and later shown to have direct relevance to human biology (Yuan et al., 1993, Cell 75:641-652; Therrien et al., 1995, Cell 83:879-888; Karim et al., 1996, Genetics 143:315-329; Kornfeld et al., 1995, Cell 83:903-913; Rubin et al., 1997, "Protein kinase required for Ras signal transduction", U.S. Patent No. 5,700,675; Steller et al., 1997, "Cell death genes of *Drosophila melanogaster* and vertebrate homologs", U.S. Patent No. 5,593,879).

SUMMARY OF THE INVENTION

The present invention relates to proteins encoded by nucleotide sequences of *D. melanogaster* insulin-like genes, as well as fragments and other derivatives and analogs of such insulin-like proteins. Nucleic acids encoding the insulin-like gene and fragments or derivatives are also within the scope of the invention. Production of the foregoing proteins, *e.g.*, by recombinant methods, is provided.

The invention also relates to insulin-like protein derivatives and analogs which are functionally active, *i.e.*, which are capable of displaying one or more known functional activities associated with a full-length (wild-type) insulin-like protein. Examples of such functional activities include antigenicity (ability to bind, or to compete for binding, to an anti-insulin-like protein antibody), immunogenicity (ability to generate antibody which binds to an insulin-like protein), and ability to bind (or

compete for binding) to a receptor for insulin (*e.g.*, that is encoded by the *D. melanogaster* insulin receptor-like gene, *InR*).

The invention further relates to fragments (and derivatives and analogs thereof) of an insulin-like protein which comprise one or more domains of the insulin-like protein.

Antibodies to an insulin-like protein, its derivatives and analogs, are additionally provided.

Methods for genetic analysis of pathways involving insulin superfamily hormones in *Drosophila* are provided. Such methods may yield results of importance to human disease. For example, systematic identification of participants in intracellular signaling by insulin-like hormones, or components regulating secretion and turnover of insulin-like hormones, provide leads to the identification of drug targets, therapeutic proteins, diagnostics, or prognostics useful for treatment or management of insulin resistance in diabetics.

BRIEF DESCRIPTION OF FIGURES

FIG. 1 illustrates the structural organization of precursor forms of the insulin superfamily of hormones. The different domains that make up precursor forms of insulin-like hormones are represented as boxes labeled Pre, F, B, C, A, D, and E. Domains that may remain in a mature hormone are represented as unshaded boxes (the B, A, and D peptide domains) or as lightly hatched (the C or "connecting" peptide domain). Domains that are removed during proteolytic processing are represented as shaded (the Pre peptide domain and F domain) or as hatched (the E peptide domain). IGF hormones are unique in having D and E peptide domains; these domains are represented as smaller boxes in FIG. 1. Cleavage sites utilized by proteases during proteolytic processing (*i.e.*, protein maturation) are indicated below the boxes. The asterisk marks the position of cleavage by signal peptidase. Arrows indicate cleavage sites by prohormone convertases. Disulfide bonds (S-S) are represented above the boxes with lines indicating connections between covalently-bonded Cys residues.

FIG. 2 illustrates conserved structural features of known insulin superfamily members. The aligned sequences of the B and A chain peptide domains are shown for representative insulin superfamily hormones from the following vertebrates and invertebrates: human insulin (Bell et al., 1979, Nature 29:525-527), human IGF-I (Jansen et al., 1983, Nature 306:609-611), human relaxin 1 (Hudson et al., 1983, Nature

301:628-631), RLF from human (Adham al., 1993, J. Biol. Chem. 268:26668-26672), placentin from human (Chassin et al., 1995, Genomics 29:465-470), bombyxin II from silkworm (Nagasawa et al., 1986, Proc. Natl. Acad. Sci. U.S.A. 83:5840-5843), MIP from freshwater snail (Smit et al., 1988, Nature 331:535-538), and LIRP from locust (Lagueux et al., 1990, Eur. J. Biochem. 187:249-254). The numbering scheme shown at the bottom of the figure is for residues of the A and B chains relative to residue numbers for human insulin peptide domains. The nearly invariant positions of the six Cys residues that participate in disulfide bonds are boxed. MIP-I is unusual in having two extra Cys residues which are also individually boxed in that sequence. Other conserved amino acid positions that play important roles in promoting the common insulin superfamily fold are highlighted by shading of the following residue positions: B6, B8, B11, B15, B18, A2, A16, and A19. Three helical regions that comprise the common insulin fold are marked above the alignments using a "<--->" symbol.

FIG. 3 shows a gene map of *Drosophila* insulin-like gene cluster region, including location and orientation of coding regions of dIns1, dIns2, dIns3, and dIns4. Units in kbp indicate kilobase pairs of genomic DNA.

FIG. 4A-4P shows the annotated genomic DNA sequence of *D. melanogaster* insulin-like gene cluster. Genomic sequence is set forth in SEQ ID NO:7.

FIG. 5 shows annotated sequence of *D. melanogaster* insulin-like protein dIns2 and corresponding cDNA. dIns2 protein sequence is set forth in SEQ ID NO:2. dIns2 nucleic acid sequence is set forth in SEQ ID NO:1.

FIG. 6 shows annotated sequence of *D. melanogaster* insulin-like protein dIns3 and corresponding cDNA. dIns3 protein sequence is set forth in SEQ ID NO:4. dIns3 nucleic acid sequence is set forth in SEQ ID NO:3.

FIG. 7 shows annotated sequence of *D. melanogaster* insulin-like protein dIns4 and corresponding cDNA. dIns4 protein sequence is set forth in SEQ ID NO:6. dIns4 nucleic acid sequence is set forth in SEQ ID NO:5.

FIG. 8 shows key structural features for *D. melanogaster* Insulin-like protein folding and conserved Cysteine residues in vertebrate superfamily. Numbers shown in parentheses represents the number of residues omitted from the C peptide sequence.

DETAILED DESCRIPTION OF THE INVENTION

Described herein are novel insulin-like genes from *Drosophila* and the characterization of their function. The *Drosophila* insulin-like genes described herein are a tightly clustered array encoding proteins that are much closer in structure to vertebrate insulins than the insulin-like proteins found in the nematode *C. elegans*. Nonetheless, the *Drosophila* insulin-like proteins exhibit significant sequence diversity. These new insulin-like genes in *Drosophila* constitute very useful tools for probing the function and regulation of their corresponding pathways. Systematic genetic analysis of signaling pathways involving insulin-like proteins in *Drosophila* can be expected to lead to the discovery of new drug targets, therapeutic proteins, diagnostics and prognostics useful in the treatment of diseases and clinical problems associated with the function of insulin superfamily hormones in humans and other animals, as well as clinical problems associated with aging and senescence. Furthermore, analysis of these same pathways using *Drosophila* insulin-like proteins as tools will have utility for identification and validation of pesticide targets in invertebrate pests that are components of these signaling pathways.

Use of *Drosophila* insulin-like genes for such purposes as disclosed herein, has advantages over manipulation of other known components of the fruit fly InR pathway including *InR, Pi3K92E*, and *chico*. First, use of ligand-encoding *Drosophila* insulin-like genes provides a superior approach for identifying factors that are upstream of the receptor in the signal transduction pathway. Specifically, components involved in the synthesis, activation and turnover of insulin-like proteins may be identified. Furthermore, the discovery of multiple, different insulin-like hormones provides a rational approach to separate components involved in responses to different, specific environmental or regulatory signals. This is less technically feasible with manipulation of downstream components of the pathway found in target tissues. Further, the diversity of different insulin-like hormones provides a means to identify potential new receptor and/or signal transduction systems for insulin superfamily hormones that are structurally different from those that have been characterized to date, in either vertebrates or invertebrates. Still further, use of *Drosophila* as a system for analyzing the function and regulation of insulin-like genes has great advantages over approaches in other organisms due to the ability to rapidly carry out large-scale, systematic genetic screens as well as the ability to screen small molecules directly on whole organisms for possible therapeutic or pesticide use. Particularly, the *Drosophila* insulin-like genes described herein are significantly

closer in structure to vertebrate insulin hormones than the insulin-like proteins of *C. elegans*; therefore, the fruit fly *Drosophila* may serve as a better model for vertebrate insulin function and signaling than the nematode *C. elegans* due to this greater structural similarity. Moreover, the fruit fly *Drosophila* is clearly the preferred genetic model organism for dissecting the function of insulin-like proteins, and validating potential pesticide targets, with respect to other insect pest species.

Isolation Of *D. Melanogaster* Insulin-Like Genes

In specific embodiments, insulin-like nucleic acids of the invention comprise the cDNA sequences of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5 or the coding regions thereof, or nucleic acids encoding an insulin-like protein (*e.g.*, a protein having the sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6). As used herein, a gene "corresponding" to a cDNA sequence shall be construed to mean the gene that encodes the RNA from which the cDNA is derived. The invention provides purified or isolated nucleic acids consisting of at least 8 nucleotides (*i.e.*, a hybridizable portion) of an insulin-like gene sequence; in other embodiments, the nucleic acids consist of at least 25 (continuous) nucleotides, 50 nucleotides, 100 nucleotides, 150 nucleotides, or 200 nucleotides of an insulin-like sequence, or a full-length insulin-like coding sequence. In another embodiment, the nucleic acids are smaller than 35, 200, or 500 nucleotides in length. Nucleic acids can be single or double stranded. The invention also relates to nucleic acids hybridizable to or complementary to the foregoing sequences or their reverse complements. In specific aspects, nucleic acids are provided which comprise a sequence complementary to at least 10, 25, 50, 100, or 200 nucleotides or the entire coding region of an insulin-like gene.

The invention further relates to the genomic nucleotide sequences of *D. melanogaster* insulin-like nucleic acids. In specific embodiments, insulin-like nucleic acids comprise the genomic sequences of SEQ ID NO:7 or the coding regions thereof, or nucleic acids encoding an insulin-like protein (*e.g.*, a protein having the sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6).

In the above or alternative embodiments, the nucleic acids of the invention consist of a nucleotide sequence of not more than 2, 5, 10, 15, or 20 kilobases.

Hybridization Conditions

A nucleic acid which is hybridizable to an insulin-like nucleic acid (*e.g.*, having a sequence as set forth in SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5, or to its reverse complement, or to a nucleic acid encoding an insulin-like derivative, or to its reverse complement), under conditions of high, medium, or low stringency is provided. Methods for selection of appropriate conditions for such stringencies is well known in the art (*see e.g.*, Sambrook et al., 1989, *supra*; Ausubel et al., eds., in the Current Protocols in Molecular Biology series of laboratory technique manuals, 8 1987-1997, Current Protocols, 8 1994-1997 John Wiley and Sons, Inc.).

An example of suitable conditions of high stringency that can be used is as follows. Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 g/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C in prehybridization mixture containing 100 g/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Washing of filters is done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA. This is followed by a wash in 0.1X SSC at 50°C for 45 min before autoradiography.

An example of procedures using conditions of medium stringency is as follows. Filters containing DNA are pretreated for 6 h at 40°C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 g/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution with the following modifications: 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 g/ml salmon sperm DNA, 10% (wt/vol) dextran sulfate, and 5-20 X 10^6 cpm ^{32}P-labeled probe is used. Filters are incubated in hybridization mixture for 18-20 h at 40°C, and then washed for 1.5 h at 55°C in a solution containing 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS. The wash solution is replaced with fresh solution and incubated an additional 1.5 h at 60°C. Filters are blotted dry and exposed for autoradiography. If necessary, filters are washed for a third time at 65-68°C and re-exposed to film.

Conditions of low stringency are as follows. Incubation for 8 hours to overnight at 37°C in a solution comprising 20% formamide, 5x SSC, 50mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 mg/ml denatured sheared salmon sperm DNA; hybridization in the same buffer for 18 to 20 hours; and washing of filters in 1x SSC at 37°C for 1 hour.

Fragments of insulin-like nucleic acids comprising regions conserved between (*i.e.*, with homology to) other insulin-like nucleic acids, of the same or different species, are also provided. Nucleic acids encoding one or more insulin-like domains are provided.

Cloning Procedures

The insulin-like genes of the invention can be cloned using any suitable technique known in the art (*see* Sambrook *et al.* (1989), *supra*; DNA Cloning: A Practical Approach, Vol. 1, 2, 3, 4, (1995) Glover, ed., MRL Press, Ltd., Oxford, U.K.). For example, with expression cloning, an expression library is constructed, mRNA is isolated, and cDNA is made and ligated into an expression vector (*e.g.*, a bacteriophage derivative) that it is capable of being expressed by the host cell into which it is then introduced. Various screening assays can then be used to select for the expressed insulin-like product such as immunoassays using anti-insulin-like antibodies.

Polymerase chain reaction (PCR) can used to amplify the desired sequence in a genomic or cDNA library, prior to selection. Oligonucleotide primers representing known insulin-like sequences can be used as primers in PCR. Preferably, the oligonucleotide primers represent at least part of conserved segments of strong homology between insulin-like genes of different species. The synthetic oligonucleotides may be utilized as primers to amplify sequences from a source (RNA or DNA), preferably a cDNA library, of potential interest. PCR can be carried out, *e.g.*, by use of a Perkin-Elmer Cetus thermal cycler and Taq polymerase (*e.g.*, Gene AmpJ). The nucleic acid being amplified can include mRNA or cDNA or genomic DNA from any species. One may synthesize degenerate primers for amplifying homologs from other species in the PCR reactions. The stringency of hybridization conditions used in priming the PCR reactions can be varied to allow for greater or lesser degrees of nucleotide sequence similarity between the known insulin-like nucleotide sequences and a nucleic acid homolog (or ortholog) being isolated. For cross species hybridization, low stringency conditions are preferred. For same species hybridization, moderately stringent conditions are preferred. After successful amplification of a segment of an insulin-like homolog, that segment may be cloned and sequenced by standard techniques, and utilized as a probe to isolate a complete cDNA or genomic clone. This, in turn, permits the determination of the gene's complete nucleotide sequence, the analysis of its expression, and the production of its protein product for functional analysis, as described below. In this fashion, additional genes encoding

insulin-like proteins and insulin-like analogs may be identified.

In another embodiment, the organizational characteristics of the insulin-like genes may be used to identify clones containing novel members of the insulin-like gene superfamily. For example, the insulin-like genes in the silkworm insect *B. Mori* (which encode the bombyxin proteins) have been demonstrated to be organized in large multi gene clusters (Kondo, et al., 1996, J. Mol. Biol. 259:926-937). Identification and characterization of the genomic region surrounding a known insulin-like gene could, therefore, be used to identify additional genes that encode insulin-like proteins or insulin-like analogs that are located within these clusters by methods known in the art.

Any eukaryotic cell potentially can serve as the nucleic acid source for molecular cloning of an insulin-like gene. The nucleic acid sequences encoding insulin-like proteins may be isolated from vertebrate, mammalian, human, porcine, bovine, feline, avian, equine, canine, as well as additional primate sources, insects (*e.g., Drosophila*), invertebrates, plants, *etc.* The DNA may be obtained by standard procedures known in the art from cloned DNA (*e.g.*, a DNA "library"), by chemical synthesis, by cDNA cloning, or by the cloning of genomic DNA, or fragments thereof, purified from the desired cell.

In the molecular cloning of the gene from genomic DNA, DNA fragments are generated, some of which will encode the desired gene. The DNA may be cleaved at specific sites using various restriction enzymes. Alternatively, one may use DNase in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The linear DNA fragments can then be separated according to size by standard techniques such as agarose and polyacrylamide gel electrophoresis and column chromatography.

Once the DNA fragments are generated, identification of the specific DNA fragment containing the desired gene may be accomplished in a number of ways. For example, if a portion of an insulin-like gene or its specific RNA or a fragment thereof is available and can be purified and labeled, the generated DNA fragments may be screened by nucleic acid hybridization to the labeled probe (Benton and Davis, 1977, Science 196:180; Grunstein and Hogness, 1975, Proc. Natl. Acad. Sci. U.S.A. 72:3961). Those DNA fragments with substantial homology to the probe will hybridize. It is also possible to identify the appropriate fragment by restriction enzyme digestion(s) and comparison of fragment sizes with those expected according to a known restriction map if such is available. Further selection can be carried out on the basis of the properties of the gene.

Alternatively, the presence of the desired gene may be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example, cDNA clones, or DNA clones which hybrid-select the proper mRNAs, can be selected and expressed to produce a protein that has, *e.g.*, similar or identical electrophoretic migration, isoelectric focusing behavior, proteolytic digestion maps, hormonal activity, binding activity, or antigenic properties as known for an insulin-like protein. Using an antibody to a known insulin-like protein, other insulin-like proteins may be identified by binding of the labeled antibody to expressed putative insulin-like proteins, *e.g.*, in an ELISA (enzyme-linked immunosorbent assay)-type procedure. Further, using a binding protein specific to a known insulin-like protein, other insulin-like proteins may be identified by binding to such a protein (*see e.g.*, Clemmons, 1993, Mol. Reprod. Dev. 35:368-374; Loddick et al., 1998, Proc. Natl. Acad. Sci. U.S.A. 95:1894-1898).

An insulin-like gene can also be identified by mRNA selection using nucleic acid hybridization followed by *in vitro* translation. In this procedure, fragments are used to isolate complementary mRNAs by hybridization. Such DNA fragments may represent available, purified insulin-like DNA of another species (*e.g.*, *Drosophila*, mouse, human). Immunoprecipitation analysis or functional assays (*e.g.*, aggregation ability *in vitro*, binding to receptor, *etc.*) of the *in vitro* translation products of the isolated products of the isolated mRNAs identifies the mRNA and, therefore, the complementary DNA fragments that contain the desired sequences. In addition, specific mRNAs may be selected by adsorption of polysomes isolated from cells to immobilized antibodies specifically directed against insulin-like protein. A radiolabeled insulin-like cDNA can be synthesized using the selected mRNA (from the adsorbed polysomes) as a template. The radiolabeled mRNA or cDNA may then be used as a probe to identify the insulin-like DNA fragments from among other genomic DNA fragments.

Alternatives to isolating the insulin-like genomic DNA include, chemically synthesizing the gene sequence itself from a known sequence or making cDNA to the mRNA which encodes the insulin-like protein. For example, RNA for cDNA cloning of the insulin-like gene can be isolated from cells which express the gene.

The identified and isolated gene can then be inserted into an appropriate cloning vector. A large number of vector-host systems known in the art may be used. Possible vectors include, plasmids or modified viruses, but the vector system must be compatible with the host cell used. Suitable vectors include bacteriophages such as lambda

derivatives, or plasmids such as PBR322 or pUC plasmid derivatives or the Bluescript vector (Stratagene USA, La Jolla, California). The insertion into a cloning vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has complementary cohesive termini. However, if the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated linkers may comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. In an alternative method, the cleaved vector and an insulin-like gene may be modified by homopolymeric tailing. Recombinant molecules can be introduced into host cells via transformation, transfection, infection, electroporation, *etc.*, so that many copies of the gene sequence are generated.

In an alternative method, the desired gene may be identified and isolated after insertion into a suitable cloning vector in a "shot gun" approach. Enrichment for the desired gene, for example, by size fractionization, can be done before insertion into the cloning vector.

In an additional embodiment, the desired gene may be identified and isolated after insertion into a suitable cloning vector using a strategy that combines a "shot gun" approach with a "directed sequencing" approach. Here, for example, the entire DNA sequence of a specific region of the genome, such as a sequence tagged site (STS), can be obtained using clones that molecularly map in and around the region of interest.

In specific embodiments, transformation of host cells with recombinant DNA molecules that incorporate an isolated insulin-like gene, cDNA, or synthesized DNA sequence enables generation of multiple copies of the gene. Thus, the gene may be obtained in large quantities by growing transformants, isolating the recombinant DNA molecules from the transformants and, when necessary, retrieving the inserted gene from the isolated recombinant DNA.

The insulin-like sequences provided by the instant invention include those nucleotide sequences encoding substantially the same amino acid sequences as found in native insulin-like proteins. and those encoded amino acid sequences with functionally equivalent amino acids, as well as those encoding other insulin-like derivatives or analogs, as described below for insulin-like derivatives and analogs.

Expression Of *D. Melanogaster* Insulin-Like Genes

The nucleotide sequence coding for an insulin-like protein or a functionally active analog or fragment or other derivative thereof, can be inserted into any appropriate expression vector that contains the necessary elements for the transcription and translation of the inserted protein-coding sequence. The necessary transcriptional and translational signals can also be supplied by the native insulin-like gene and/or its flanking regions. A variety of host-vector systems may be utilized to express the protein-coding sequence such as mammalian cell systems infected with virus (*e.g.*, vaccinia virus, adenovirus, *etc.*); insect cell systems infected with virus (*e.g.*, baculovirus); microorganisms such as yeast containing yeast vectors, or bacteria transformed with bacteriophage, DNA, plasmid DNA, or cosmid DNA. The expression elements of vectors vary in their strengths and specificities. Depending on the host-vector system utilized, any one of a number of suitable transcription and translation elements may be used. In yet another embodiment, a fragment of an insulin-like protein comprising one or more domains of the insulin-like protein is expressed.

Any of the methods previously described for the insertion of DNA fragments into a vector may be used to construct expression vectors containing a chimeric gene consisting of appropriate transcriptional/translational control signals and the protein coding sequences. These methods may include *in vitro* recombinant DNA and synthetic techniques and *in vivo* recombinants (genetic recombination). Expression of a nucleic acid sequence encoding an insulin-like protein or peptide fragment may be regulated by a second nucleic acid sequence so that the insulin-like protein or peptide is expressed in a host transformed with the recombinant DNA molecule. For example, expression of an insulin-like protein may be controlled by any promoter/enhancer element known in the art including those of prokaryotic expression vectors and plant expression vectors; promoter elements from yeast or other fungi; and transcriptional control regions. In some embodiments, the promoter will exhibit tissue specificity. In a specific embodiment, a vector is used that comprises a promoter operably linked to an insulin-like gene nucleic acid, one or more origins of replication, and, optionally, one or more selectable markers (*e.g.*, an antibiotic resistance gene).

Expression constructs can be made by subcloning an insulin-like coding sequence into the EcoRI restriction site of each of the three pGEX vectors (Smith and Johnson, 1988, Gene 7:31-40). This allows for the expression of the insulin-like protein

product from the subclone in the correct reading frame. Expression vectors containing insulin-like gene inserts can be identified by three general approaches: (a) nucleic acid hybridization; (b) presence or absence of "marker" gene functions; and (c) expression of inserted sequences. In the first approach, the presence of an insulin-like gene inserted in an expression vector can be detected by nucleic acid hybridization using probes comprising sequences that are homologous to an inserted insulin-like gene. In the second approach, the recombinant vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene functions (*e.g.*, thymidine kinase activity, resistance to antibiotics, transformation phenotype, occlusion body formation in baculovirus, *etc.*) caused by the insertion of an insulin-like gene in the vector. For example, if the insulin-like gene is inserted within the marker gene sequence of the vector, recombinants containing the insulin-like insert can be identified by the absence of the marker gene function. In the third approach, recombinant expression vectors can be identified by assaying the insulin-like product expressed by the recombinant. Such assays can be based, for example, on the physical or functional properties of the insulin-like protein in *in vitro* assay systems, *e.g.*, binding with anti-insulin-like protein antibody.

Once a particular recombinant DNA molecule is identified and isolated, several methods known in the art may be used to propagate it. Once a suitable host system and growth conditions are established, recombinant expression vectors can be propagated and prepared in quantity. Some of the expression vectors which can be used include human or animal viruses such as vaccinia virus or adenovirus; insect viruses such as baculovirus; yeast vectors; bacteriophage vectors (*e.g.*, lambda phage), and plasmid and cosmid DNA vectors.

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Expression from certain promoters can be elevated in the presence of certain inducers; thus, expression of the genetically engineered insulin-like protein may be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (*e.g.*, glycosylation, phosphorylation of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. For example, expression in a bacterial system can be used to produce a non-

glycosylated core protein product. Expression in yeast will produce a glycosylated product. Expression in mammalian cells can be used to ensure "native" glycosylation of a heterologous protein. Furthermore, different vector/host expression systems may effect processing reactions to different extents.

In other embodiments of the invention, the insulin-like protein, fragment, analog, or derivative may be expressed as a fusion, or chimeric protein product (comprising the protein, fragment, analog, or derivative joined via a peptide bond to a heterologous protein sequence of a different protein). Such a chimeric product can be made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other by methods known in the art, in the proper coding frame, and expressing the chimeric product by methods commonly known in the art. Alternatively, such a chimeric product may be made by protein synthetic techniques, *e.g.*, by use of a peptide synthesizer.

Identification And Purification Of Gene Products

In particular aspects, the invention provides amino acid sequences of insulin-like proteins and fragments and derivatives thereof which comprise an antigenic determinant (*i.e.*, can be recognized by an antibody) or which are otherwise functionally active, as well as nucleic acid sequences encoding the foregoing. "Functionally active" insulin-like material as used herein refers to that material displaying one or more functional activities associated with a full-length (wild-type) insulin-like protein, *e.g.*, binding to an insulin-like receptor (*e.g.*, InR or insulin-like protein binding partner, antigenicity (binding to an anti- insulin-like protein antibody), immunogenicity, *etc.*

In specific embodiments, the invention provides fragments of an insulin-like protein consisting of at least 10 amino acids, 20 amino acids, 50 amino acids, or of at least 75 amino acids. In other embodiments, the proteins comprise or consist essentially of an insulin-like B peptide domain, an insulin-like A peptide domain, an insulin-like C peptide domain, or any combination of the foregoing, of an insulin-like protein. Fragments, or proteins comprising fragments, lacking some or all of the foregoing regions of a insulin-like protein are also provided. Nucleic acids encoding the foregoing are provided. In specific embodiments, the foregoing proteins or fragments are not more than 25, 50, or 100 contiguous amino acids.

Once a recombinant which expresses the insulin-like gene sequence is identified,

the gene product can be analyzed. This is achieved by assays based on the physical or functional properties of the product, including radioactive labeling of the product followed by analysis by gel electrophoresis, immunoassay, *etc.*

Once the insulin-like protein is identified, it may be isolated and purified by standard methods including chromatography (*e.g.*, ion exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. The functional properties may be evaluated using any suitable assay.

Alternatively, once an insulin-like protein produced by a recombinant is identified, the amino acid sequence of the protein can be deduced from the nucleotide sequence of the chimeric gene contained in the recombinant. As a result, the protein can be synthesized by standard chemical methods known in the art (*e.g.*, *see* Hunkapiller et al., 1984, Nature 310:105-111).

In another alternate embodiment, native insulin-like proteins can be purified from natural sources, by standard methods such as those described above (*e.g.*, immunoaffinity purification).

In a specific embodiment of the present invention, such insulin-like proteins, whether produced by recombinant DNA techniques or by chemical synthetic methods or by purification of native proteins, include but are not limited to those containing, as a primary amino acid sequence, all or part of the amino acid sequence substantially as depicted in Figs. 5-7 (SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, respectively), as well as fragments and other derivatives, and analogs thereof, including proteins homologous thereto.

Structure Of Insulin-Like Genes And Proteins

The structure of insulin-like genes and proteins of the invention can be analyzed by various methods known in the art, including genetic analysis and protein analysis.

Genetic analysis methods for determining the structure of cloned DNA or cDNA corresponding to an insulin-like include Southern hybridization, Northern hybridization, restriction endonuclease mapping, and DNA sequence analysis. Accordingly, this invention provides nucleic acid probes recognizing an insulin-like gene. For example, polymerase chain reaction followed by Southern hybridization with an insulin-like gene-specific probe can allow the detection of an insulin-like gene in DNA from various

cell types. Methods of amplification other than PCR are commonly known and can also be employed. In one embodiment, Southern hybridization can be used to determine the genetic linkage of an insulin-like gene. Northern hybridization analysis can be used to determine the expression of an insulin-like gene. Various cell types, at various states of development or activity can be tested for insulin-like gene expression. The stringency of the hybridization conditions for both Southern and Northern hybridization can be manipulated to ensure detection of nucleic acids with the desired degree of relatedness to the specific insulin-like gene probe used. Modifications of these methods and other methods commonly known in the art can be used.

Restriction endonuclease mapping can be used to roughly determine the genetic structure of an insulin-like gene. Restriction maps derived by restriction endonuclease cleavage can be confirmed by DNA sequence analysis.

DNA sequence analysis can be performed by any techniques known in the art, such as the method of Maxam and Gilbert (1980, Meth. Enzymol. 65:499-560), the Sanger dideoxy method (Sanger et al., 1977, Proc. Natl. Acad. Sci. U.S.A. 74:5463), the use of T7 DNA polymerase (Tabor and Richardson, U.S. Patent No. 4,795,699), or use of an automated DNA sequenator (*e.g.*, Applied Biosystems, Foster City, California).

The amino acid sequence of an insulin-like protein can be derived by deduction from the DNA sequence, or alternatively, by direct sequencing of the protein, *e.g.*, with an automated amino acid sequencer. An insulin-like protein sequence can be further characterized by a hydrophilicity analysis (Hopp and Woods, 1981, Proc. Natl. Acad. Sci. U.S.A. 78:3824). A hydrophilicity profile can be used to identify the hydrophobic and hydrophilic regions of the insulin-like protein and the corresponding regions of the gene sequence that encode such regions.

Secondary, structural analysis (Chou and Fasman, 1974, Biochemistry 13:222) can also be done, to identify regions of an insulin-like protein that assume specific secondary structures.

Manipulation, translation, and secondary structure prediction, open reading frame prediction and plotting, as well as determination of sequence homologies, can also be accomplished using computer software programs available in the art.

Other methods of structural analysis include X-ray crystallography, nuclear magnetic resonance spectroscopy and computer modeling.

Antibodies

The insulin-like protein of SEQ ID NOs:2, 4 and 6, or fragments or derivatives thereof, may be used as an immunogen to generate antibodies. Such antibodies include polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library. In another embodiment, antibodies to a domain (*e.g.*, an insulin-like receptor binding domain) of an insulin-like protein are produced. In a specific embodiment, fragments of an insulin-like protein identified as hydrophilic are used as immunogens for antibody production using art-known methods. Some examples of suitable techniques include methods which provides for the production of antibody molecules by continuous cell lines in culture; the production of monoclonal antibodies in germ-free animals (*see e.g.*, PCT/US90/02545); the use of human hybridomas (Cole et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:2026-2030); transforming human B cells with EBV virus *in vitro* (Cole et al., 1985, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, pp. 77-96). Additionally, known techniques can be used for the production of "chimeric antibodies" (e.g. by splicing the genes from a mouse antibody molecule specific for an insulin-like protein together with genes from a human antibody molecule of appropriate biological activity), insulin-like-specific single chain antibodies; and Fab expression libraries (e.g. to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for insulin-like proteins, derivatives, or analogs). The foregoing antibodies can be used against the insulin-like protein sequences described herein, *e.g.*, for imaging these proteins, measuring levels thereof, in diagnostic methods, *etc.*

Insulin-Like Proteins, Derivatives And Analogs

The invention relates to insulin-like proteins and derivatives, fragments, and analogs thereof, as well as the nucleic acids encoding them. In one embodiment, the insulin-like proteins are encoded by the insulin-like nucleic acids described above. In particular aspects, the proteins, derivatives, or analogs are of insulin-like proteins encoded by the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:7.

In a specific embodiment, the insulin-like protein fragment, derivative or analog is functionally active, *i.e.*, capable of exhibiting one or more functional activities associated with a full-length, wild-type insulin-like protein. As one example, such fragments, derivatives or analogs have the desired immunogenicity or antigenicity for use in immunoassays, for immunization, for inhibition of insulin-like activity, *etc.* As another

example, such fragments, derivatives or analogs which have the desired binding activity can be used for binding to the InR gene product. As yet another example, they have the desired binding activity can be used for binding to a binding protein specific for a known insulin-like protein (*see e.g.*, Clemmons, 1993, Mol. Reprod. Dev. 35:368-374; Loddick et al., 1998, Proc. Natl. Acad. Sci. U.S.A. 95:1894-1898). Derivatives or analogs that retain, or alternatively lack or inhibit, a desired insulin-like protein property-of-interest (*e.g.*, binding to an insulin-like protein binding partner), can be used as inducers, or inhibitors, respectively, of such property and its physiological correlates. A specific embodiment relates to an insulin-like protein fragment that can be bound by an anti-insulin-like protein antibody. Derivatives or analogs of an insulin-like protein can be tested for the desired activity by procedures known in the art.

In particular, insulin-like derivatives can be made by altering insulin-like sequences by substitutions, additions (*e.g.*, insertions) or deletions that provide for functionally equivalent molecules. Due to the degeneracy of nucleotide coding sequences, other DNA sequences which encode substantially the same amino acid sequence as an insulin-like gene may be used in the practice of the present invention. These include nucleotide sequences comprising all or portions of an insulin-like gene which is altered by the substitution of different codons that encode a functionally equivalent amino acid residue within the sequence, thus producing a silent change. Likewise, the insulin-like derivatives of the invention include those containing, as a primary amino acid sequence, all or part of the amino acid sequence of an insulin-like protein including altered sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence resulting in a silent change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent alteration. Substitutions for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Such substitutions are generally understood to be conservative substitutions.

In a specific embodiment of the invention, proteins consisting of or comprising a fragment of an insulin-like protein of at least 10 (continuous) amino acids of the insulin-like protein is provided. In other embodiments, the fragment comprises at least 20, 50, or 75 amino acids of the insulin-like protein. In specific embodiments, such fragments are not larger than 35, 100 or 200 amino acids. Derivatives or analogs of insulin-like proteins may comprise regions that are substantially homologous to an insulin-like protein or fragment thereof (*e.g.*, in various embodiments, at least 60% or 70% or 80% or 90% or 95% identity over an amino acid sequence of identical size. As used herein, "percent (%) sequence identity" with respect to a subject sequence, or a specified portion of a subject sequence, is defined as the percentage of nucleotides or amino acids in the candidate sequence identical with the nucleotides in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul *et al.*, J. Mol. Biol. (1997) 215:403-410; http://blast.wustl.edu/blast/README.html; (hereinafter referred to generally as "BLAST") with all the search parameters set to default values. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched. A percent (%) identity value is determined by the number of matching identical nucleotides or amino acids divided by the sequence length for which the percent identity is being reported.

The insulin-like derivatives and analogs of the invention can be produced by various methods known in the art. The manipulations which result in their production can occur at the gene or protein level. For example, a cloned insulin-like gene sequence can be modified by any of numerous strategies known in the art (Sambrook et al., 1989). The sequence can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*. Additionally, an insulin-like nucleic acid sequence can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or to form new restriction endonuclease sites or destroy preexisting ones, to facilitate further *in vitro* modification. Any technique for mutagenesis known in the art can be used, for example, chemical mutagenesis, *in vitro* site-directed mutagenesis (Hutchinson et al., 1978, J. Biol. Chem. 253:6551), use of TAB7 linkers (Pharmacia),

PCR with primers containing a mutation, *etc.*

Manipulations of an insulin-like protein sequence may also be made at the protein level. Included within the scope of the invention are insulin-like protein fragments or other derivatives or analogs which are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, *etc.* Any of numerous chemical modifications may be carried out by known techniques, such as specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, $NaBH_4$, acetylation, formylation, oxidation, reduction, metabolic synthesis in the presence of tunicamycin, *etc.*

In addition, analogs and derivatives of an insulin-like protein can be chemically synthesized. For example, a peptide corresponding to a portion of an insulin-like protein which comprises the desired domain, or which mediates the desired activity *in vitro*, can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the insulin-like sequence. Examples of non-classical amino acids include the D-isomers of the common amino acids, -amino isobutyric acid, 4- aminobutyric acid, Abu, 2-amino butyric acid, -Abu, -Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, -alanine, fluoro-amino acids, designer amino acids such as -methyl amino acids, C-methyl amino acids, N-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

In a specific embodiment, an insulin-like protein derivative is a chimeric or fusion protein comprising an insulin-like protein or fragment thereof (preferably consisting of at least a domain or motif of the insulin-like protein, or at least 10 amino acids of the insulin-like protein) joined at its amino- or carboxy-terminus via a peptide bond to an amino acid sequence of a different protein. In specific embodiments, the amino acid sequence of the different protein is at least 6, 10, 20 or 30 continuous amino acids of the different proteins or a portion of the different protein that is functionally active. In one embodiment, such a chimeric protein is produced by recombinant expression of a nucleic acid encoding the protein (comprising an insulin-like-coding sequence joined in-frame to a coding sequence for a different protein). Such a chimeric product can be made by ligating

the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other by methods known in the art, in the proper coding frame, and expressing the chimeric product by methods commonly known in the art. Alternatively, such a chimeric product may be made by protein synthetic techniques, *e.g.*, by use of a peptide synthesizer. Chimeric genes comprising portions of an insulin-like gene fused to any heterologous protein-encoding sequences may be constructed. A specific embodiment relates to a chimeric protein comprising a fragment of an insulin-like protein of at least six amino acids, or a fragment that displays one or more functional activities of the insulin-like protein.

In another specific embodiment, the insulin-like derivative is a molecule comprising a region of homology with a insulin-like protein. By way of example, in various embodiments, a first protein region can be considered "homologous" to a second protein region when the amino acid sequence of the first region is at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, or 95% identical, when compared to any sequence in the second region of an equal number of amino acids as the number contained in the first region or when compared to an aligned sequence of the second region that has been aligned by a computer homology program known in the art. For example, a molecule can comprise one or more regions homologous to an insulin-like domain or a portion thereof.

In a specific embodiment, the invention relates to insulin-like derivatives and analogs, in particular insulin-like fragments and derivatives of such fragments, that comprise, or alternatively consist of, one or more domains of an insulin-like protein, including but not limited to an insulin-like B peptide domain, an insulin-like A peptide domain, or an insulin-like connecting (C) peptide domain.

A specific embodiment relates to molecules comprising specific fragments of an insulin-like protein that are those fragments in the respective insulin-like proteins of the invention most homologous to specific fragments of a human or mouse insulin-like protein. A fragment comprising a domain of an insulin-like homolog can be identified by protein analysis methods well known in the art. In another specific embodiment, a molecule is provided that comprises one or more domains (or functional portion thereof) of an insulin-like protein. In particular examples, insulin-like protein derivatives are provided that contain either an A peptide domain or a B peptide domain. By way of another example, such a protein may retain such domains separated by a peptide spacer. Such spacer may be the same as or different from an insulin-like connecting (C) peptide.

In another embodiment, a molecule is provided that comprises one or more domains (or functional portion(s) thereof) of an insulin-like protein, and that has one or more mutant (*e.g.*, due to deletion or point mutation(s)) domains of an insulin-like protein (*e.g.*, such that the mutant domain has decreased function).

5

Generation And Genetic Analysis Of *Drosophila* With Altered Insulin-Like Genes

The present invention provides for methods of creating genetically-engineered fruit flies and laboratory-generated mutant fruit flies.

Genetically-engineered fruit flies can be made that harbor one or more deletions or 10 insertions in an insulin-like gene or genes. In another embodiment, genetically-engineered fruit flies harbor interfering RNAs derived from such genes. In another embodiment, genetically-engineered fruit flies harbor one or more transgenes for mis-expression of wild-type or mutant forms of such genes. Mutant fruit flies can be generated in the lab to contain deletions, insertions, rearrangements, or point mutations in an insulin-like gene or 15 genes, or combinations thereof.

The present invention provides a method by which *Drosophila* strains with laboratory-generated alterations in insulin-like genes may be used for the identification of insulin-like genes that participate in particular biochemical and/or genetic pathways. In a specific embodiment, *Drosophila* strains with laboratory-generated alterations in one or 20 more insulin-like genes may be used for the identification of insulin-like genes that participate in biochemical and/or genetic pathways that constitute possible pesticide targets, as judged by phenotypes such as non-viability, block of normal development, defective feeding, defective movement, or defective reproduction. That is, development of such a phenotype in a *Drosophila* containing an alteration in a *Drosophila* insulin-like 25 gene indicates that the insulin-like gene is a potential pesticide target.

In another embodiment, *Drosophila* strains with laboratory-generated alterations relate to therapeutic applications associated with the insulin superfamily hormones, such as metabolic control, growth regulation, differentiation, reproduction, and aging.

In another embodiment, *Drosophila* strains with laboratory-generated alterations 30 relate to large-scale genetic modifier screens aimed at systematic identification of components of genetic and/or biochemical pathways that serve as novel drug targets, diagnostics, prognostics, therapeutic proteins, pesticide targets or protein pesticides.

The invention provides methods for creating and analyzing *Drosophila* strains

having modified expression of insulin-like genes, as described below. In one embodiment, expression modification methods include any method known to one skilled in the art. Specific examples include chemical mutagenesis, transposon mutagenesis, antisense RNA interference, and transgene-mediated mis-expression. In the creation of transgenic animals, it is preferred that heterologous (*i.e.*, non- native) promoters be used to drive transgene expression.

Generation Of Loss-Of-Function Mutation In Insulin-Like Gene

The present invention provides methods of testing for preexisting mutations in a *D. melanogaster* insulin-like gene. In a specific embodiment, the genomic sequence containing the entire insulin cluster can be used to determine whether an existing mutant *Drosophila* line corresponds to a mutation in one or more of the insulin-like genes. Mutations in genes that map to the same genetic region as the insulin-like gene cluster (chromosomal band 67C-D) are of particular interest. For example, a large number of previously identified mutations have been mapped to the approximate genetic region of the insulin cluster (67C-D), including l(3)67BDa, l(3)67BDb, l(3)67BDc, l(3)67BDd, l(3)67BDe, l(3)67BDf, l(3)67BDg l(3)67BDh, l(3)67BDi l(3)67BDj, l(3)67BDk, l(3)67BDl, l(3)67BDm, l(3)67BDn, l(3)67BDp, l(3)67BDq, l(3)67BDr (FlyBase: a *Drosophila* database, Flybase consortium, Harvard University); however, the normal function of these genes has not been determined. To ascertain whether any of these mutations are in an insulin-like gene, a genomic fragment containing the *Drosophila* insulin gene cluster and potential flanking regulatory regions can be subcloned into any appropriate *Drosophila* transformation vector, such as the Carnegie series of vectors (Rubin and Spradling, 1983, Nucleic Acids Res. 11(18):6341-51), the pCaspeR series of vectors (Thummel, et al., 1988, Gene 74(2):445-56), or the pW8 vector (Klemenz, et al., 1987, Nucleic Acids Res. 15(10):3947-59) and injected into flies along with an appropriate helper plasmid to supply transposase. Resulting transformants are crossed for complementation testing to an existing panel of *Drosophila* lines containing mutations that have been mapped to the appropriate genomic region (67C-D) as described above (Greenspan, 1997, in *Fly pushing: The Theory and Practice of Drosophila Genetics* Cold Spring Harbor Press, Plainview, NY, pp. 3-46). If a mutant line is discovered to be rescued by this genomic fragment, as judged by complementation of the mutant phenotype, progressively smaller subclones or clones containing a single insulin gene can

be individually tested until the responsible locus is identified.

Generating Loss-Of-Function Mutations By Mutagenesis

Further, the invention herein provides a method for generating loss-of-function mutations in a *D. melanogaster* insulin-like gene. Mutations can be generated by one of many mutagenesis methods known to investigators skilled in the art (Ashburner, 1989, In *Drosophila: A Laboratory Manual*, Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: pp. 299-418.; "*Fly pushing: The Theory and Practice of Drosophila Genetics*" Cold Spring Harbor Press, Plainview, NY). In a specific embodiment, the mutagens that can be used include but are not restricted to: transposons such as the P or hobo elements; chemical mutagens such as ethylmethane sulfonate (EMS), methylmethane sulfonate (MMS), N-ethyl-N-nitrosourea (ENU), triethylmelamine, diepoxyalkanes, ICR-170, or formaldehyde; and irradiation with X-rays, gamma rays, or ultraviolet radiation.

Mutagenesis by P elements, or marked P elements, is particularly appropriate for isolation of loss-of-function mutations in *Drosophila* insulin-like genes due to the precise molecular mapping of these genes, the small size of these targets, the availability and proximity of preexisting P element insertions for use as a localized transposon source, and the potential to knock out several of these genes by induction of a small deletion of the locus (Hamilton and Zinn, 1994, Methods in Cell Biology 44:81-94; Wolfner and Goldberg, 1994, Methods in Cell Biology 44:33-80; Clark, et al., 1994, Proc. Natl. Acad. Sci. U.S.A. 91(2):719-22; Kaiser, 1990, Bioessays 12(6):297-301, In *Drosophila melanogaster: Practical Uses in Cell and Molecular Biology*, L.S.B. Goldstein and E. A. Fyrberg, Eds., Academic Press, Inc. San Diego, California). For the purposes of mutagenesis, modified P elements are typically used which contain one or more of the following elements: sequences encoding a dominant visible marker, usually a wild-type *white+* or *rosy+* eye color gene, to allow detection of animals containing the P element and to screen for transposition events (Rubin and Spradling, 1982, Science 218(4570):348-53; Klemenz, et al., 1987, Nucleic Acids Res. 15(10):3947-59), bacterial plasmid sequences including a selectable marker such as ampicillin resistance to facilitate cloning of genomic sequences adjacent to the insertion site (Steller and Pirrotta, 1985, Embo. J. 4:167-171) and *lacZ* sequences fused to a weak general promoter to detect the presence of enhancers with a developmental expression pattern of interest (Bellen, et al., 1989, Genes Dev.

3(9):1288-300; Bier, et al., 1989, Genes Dev. 3(9):1273-87; Wilson, et al., 1989, Genes Dev. 3(9):1301-13). For examples of marked P elements useful for mutagenesis *see* "FlyBase - A Drosophila Database", Nucleic Acids Research 26:85-88, (http://flybase.bio.indiana.edu).

A preferred method of transposon mutagenesis employs the "local hopping" method (Tower et al., 1993, Genetics 133:347-359). Briefly, an existing mutant *Drosophila* line containing a P element inserted into chromosomal bands 67C-D, such as l(3)01859 or any other P element that maps within this region, is crossed to a *Drosophila* line expressing transposase in order to mobilize the transposon. Transposition of the P element, which contains a marker gene that typically affects eye color, is determined phenotypically on the basis of eye color change in the resulting progeny. Candidate insertion lines are selected for further analysis on the basis of close linkage of the new insertion to the initial insertion site, which can be determined by standard genetic mapping techniques such as high frequency cosegregation of markers. Each new P insertion line can be tested molecularly for transposition of the P element into the insulin-like gene cluster by assays based on PCR amplification. For each reaction, one PCR primer is used that is homologous to sequences contained within the P element and a second primer is homologous to one of the individual insulin genes, in either the coding region or flanking regions of the insulin-like gene. Products of the PCR reactions are detected by agarose gel electrophoresis. The sizes of the resulting DNA fragments are used to map the site of P element insertion.

Alternatively, Southern blotting and restriction mapping using DNA probes derived from genomic DNA or cDNAs of the insulin-like genes can be used to detect transposition events that rearrange the genomic DNA of the insulin-like genes. P transposition events that map to the insulin gene cluster can be assessed for phenotypic effects in heterozygous or homozygous mutant *Drosophila*, as described in detail below.

Generating Localized Deletions In The Insulin Gene Cluster

In another embodiment, *Drosophila* lines carrying P insertions in the insulin gene cluster can be used to generate localized deletions in the insulin-like gene cluster by previously described methods known in the art (Kaiser, 1990, Bioessays 12(6):297-301; Harnessing the power of Drosophila genetics, In *Drosophila melanogaster: Practical Uses in Cell and Molecular Biology*, L.S.B. Goldstein and E. A. Fyrberb, eds., Academic Press,

Inc. San Diego, California). This is particularly useful if no P elements transpositions are found that disrupt a particular insulin-like gene of interest. In brief, flies containing P elements inserted into the insulin gene cluster are exposed to a further round of transposase to induce excision of the element. Progeny in which the transposon has excised are typically identified by loss of the eye color marker associated with the transposable element. The resulting progeny will include flies with either precise or imprecise excision of the P element, where the imprecise excision events often result in deletion of genomic DNA neighboring the site of P insertion. Such progeny can be screened by molecular techniques to identify deletion events that remove flanking genomic sequence. Such methods include, (a) methods of detecting alterations in the genomic DNA based on PCR amplification with primers flanking the insertion site of the P element; (b) methods based on Southern blotting and restriction mapping using DNA probes derived from the P element, DNA probes derived from flanking genomic sequence in the region of the insulin-like genes, or DNA probes derived from cDNAs of insulin-like genes. Deletions generated in this manner that remove one or more of the insulin-like loci can be assessed for phenotypic effects in heterozygous and homozygous mutant *Drosophila* as described below.

Generating Loss-Of-Function Phenotypes Using Methods Based On RNA-Mediated Interference With Gene Expression

The invention further provides a method for generating loss-of-function phenotypes using methods based on RNA-mediated interference with gene expression. The function of the *Drosophila* insulin-like genes identified herein may be characterized and/or determined by generating loss-of-function phenotypes through such RNA-based methods.

In one embodiment, loss-of-function phenotypes are generated by antisense RNA methods (Schubiger and Edgar, 1994, Methods in Cell Biology 44:697-713). One form of the antisense RNA method involves the injection of embryos with an antisense RNA that is partially homologous to the gene-of-interest (in this case an insulin-like gene). Another form of the antisense RNA method involves expression of an antisense RNA partially homologous to the gene-of-interest by operably joining a portion of the gene-of-interest in the antisense orientation to a powerful promoter that can drive the expression of large quantities of antisense RNA, either generally throughout the animal or in specific tissues.

Examples of powerful promoters that can be used in this strategy of antisense RNA include heat shock gene promoters or promoters controlled by potent exogenous transcription factors, such as GAL4 and tTA, described in more detail in the following section. Antisense RNA-generated loss-of-function phenotypes have been reported previously for several *Drosophila* genes including *cactus*, *pecanex*, and *Krupple* (LaBonne, et al., 1989, Dev. Biol. 136(1):1-16; Schuh and Jackle, 1989, Genome 31(1):422-5; Geisler, et al., 1992, Cell 71(4):613-21).

In a second embodiment, loss-of-function phenotypes are generated by cosuppression methods (Bingham, 1997, Cell 90(3):385-7; Smyth, 1997, Curr. Biol. 7(12):793-5; Que and Jorgensen, 1998, Dev. Genet. 22(1):100-9). Cosuppression is a phenomenon of reduced gene expression produced by expression or injection of a sense strand RNA corresponding to a partial segment of the gene-of-interest. Cosuppression effects have been employed extensively in plants to generate loss-of-function phenotypes, and there is report of cosuppression in *Drosophila* where reduced expression of the *Adh* gene was induced from a *white-Adh* transgene (Pal-Bhadra, et al., 1997, Cell 90(3):479-90).

In a third embodiment, loss-of-function phenotypes may be generated by double-stranded RNA interference. This method is based on the interfering properties of double-stranded RNA derived from the coding regions of genes. Termed dsRNAi, this method has proven to be of great utility in genetic studies of the nematode *C. elegans* (*see* Fire et al., 1998, Nature 391:806-811). In a preferred embodiment of this method, complementary sense and antisense RNAs derived from a substantial portion of a gene-of-interest, such as an insulin-like gene, are synthesized *in vitro*. Phagemid DNA templates containing cDNA clones of the gene-of-interest are inserted between opposing promoters for T3 and T7 phage RNA polymerases. Alternatively, one can use PCR products amplified from coding regions of insulin-like genes, where the primers used for the PCR reactions are modified by the addition of phage T3 and T7 promoters. The resulting sense and antisense RNAs are annealed in an injection buffer, and the double-stranded RNA injected or otherwise introduced into animals. Progeny of the injected animals are then inspected for phenotypes-of-interest.

ANTISENSE REGULATION OF GENE EXPRESSION

The invention provides for antisense uses of *D. melanogaster* insulin-like genes.

In a specific embodiment, an insulin-like protein function is inhibited by use of insulin-like antisense nucleic acids. The present invention provides for use of nucleic acids of at least six nucleotides that are antisense to a gene or cDNA encoding an insulin-like protein or a portion thereof. An insulin-like "antisense" nucleic acid as used herein refers to a nucleic acid capable of hybridizing to a sequence-specific (i.e. non-poly A) portion of an insulin-like RNA (preferably mRNA) by virtue of some sequence complementarily. Antisense nucleic acids may also be referred to as inverse complement nucleic acids. The antisense nucleic acid may be complementary to a coding and/or noncoding region of an insulin-like mRNA. Such antisense nucleic acids have utility in inhibiting an insulin-like protein function. For example, such antisense nucleic acids may be useful as pesticides to eradicate parasites in plants, or in animals such as dogs, horses, and cattle.

The antisense nucleic acids of the invention can be oligonucleotides that are double-stranded or single-stranded, RNA or DNA or a modification or derivative thereof, which can be directly administered to a cell, or which can be produced intracellularly by transcription of exogenous introduced sequences. In a preferred embodiment, the antisense nucleic acids of the invention are double-stranded RNA mentioned previously (*see* Fire et al., 1998, Nature 391:806-811).

The insulin-like antisense nucleic acids of the invention are preferably oligonucleotides (ranging from 6 to about 50 oligonucleotides). In specific aspects, an oligonucleotide is at least 10 nucleotides, at least 15 nucleotides, at least 100 nucleotides, or at least 200 nucleotides in length. The oligonucleotide can be DNA or RNA or chimeric mixtures or derivatives or modified versions thereof, or single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone. The oligonucleotide may include other appending groups such as peptides, or agents facilitating transport across the cell membrane (*see e.g.*, Letsinger et al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86:6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. U.S.A. 84:648-652; PCT Publication No. WO 88/09810, published December 15, 1988) or the blood-brain barrier (*see e.g.*, PCT Publication No. WO 89/10134, published April 25, 1988), hybridization-triggered cleavage agents (*see e.g.*, Krol et al., 1988, BioTechniques 6:958-976) or intercalating agents (*see e.g.*, Zon, 1988, Pharm. Res. 5:539-549).

In a preferred aspect of the invention, an insulin-like antisense oligonucleotide is

provided as single-stranded DNA. In another preferred aspect, such an oligonucleotide comprises a sequence antisense to the sequence encoding a B peptide domain or an A peptide domain of an insulin-like protein. The oligonucleotide may be modified at any position on its structure with substituents generally known in the art.

The insulin-like antisense oligonucleotide may comprise at least one modified base moiety for example, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5.-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. In another embodiment, the oligonucleotide comprises at least one modified sugar moiety, for example, arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the oligonucleotide comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

In yet another embodiment, the oligonucleotide is an -anomeric oligonucleotide. An -anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual -units, the strands run parallel to each other (Gautier et al., 1987, Nucl. Acids Res. 15:6625-6641). The oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization-triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, *etc.*

Oligonucleotides of the invention may be synthesized by standard methods known in the art, *e.g.*, by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, *etc.*). As examples, phosphorothioate oligonucleotides may be synthesized by the method of Stein et al. (Stein et al., 1988, Nucl.

Acids Res. 16:3209), methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451), *etc.*

In a specific embodiment, an insulin-like antisense oligonucleotide comprises catalytic RNA, or a ribozyme (*see e.g.*, PCT Publication WO 90/11364, published October 4, 1990; Sarver et al., 1990, Science 247:1222-1225). In another embodiment, the oligonucleotide is a 2.-0-methylribonucleotide (Inoue et al., 1987, Nucl. Acids Res. 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330).

In an alternative embodiment, the insulin-like antisense nucleic acid of the invention is produced intracellularly by transcription from an exogenous sequence. For example, a vector can be introduced *in vivo* such that it is taken up by a cell, within which cell the vector or a portion thereof is transcribed, producing an antisense nucleic acid (RNA) of the invention. Such a vector would contain a sequence encoding the insulin-like antisense nucleic acid. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. Expression of the sequence encoding the insulin-like antisense RNA can be by any promoter known in the art. Such promoters can be inducible or constitutive. Examples include the SV40 early promoter region (Benoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3. long terminal repeat of Rous sarcoma virus (Yamamoto et al., 1980, Cell 22:787-797), the herpes thymidine kinase promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. U.S.A. 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296:39-42), *etc.*

The antisense nucleic acids of the invention comprise a sequence complementary to at least a sequence-specific portion of an RNA transcript of an insulin-like gene. However, absolute complementarity, although preferred, is not required. A sequence "complementary to at least a portion of an RNA," as referred to herein, means a sequence having sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded insulin-like antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to

hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an insulin-like RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine, e.g., the melting point of the hybridized complex.

Generating Gain-Of-Function Phenotypes By Ectopic Expression Of Insulin-Like Genes

The current invention provides methods for generating gain-of-function phenotypes by ectopic expression of insulin-like genes. Ectopic expression, including mis-expression or overexpression, of wild type or altered *Drosophila* insulin-like genes in transgenic animals is another useful method for the analysis of gene function (Brand, et al., 1994, Methods in Cell Biology 44:635-654, Ectopic expression in *Drosophila*; Hay, et al., 1997, Proc. Natl. Acad. Sci. U.S.A. 94(10):5195-200). Such transgenic *Drosophila* may be created that contain gene fusions of the coding regions of insulin-like genes (from either genomic DNA or cDNA) operably joined to a specific promoter and transcriptional enhancer whose regulation has preferably been well characterized, preferably heterologous promoters/enhancers that do not normally drive the expression of the insulin-like genes. Examples of promoters/enhancers that can be used to drive such misexpression of insulin-like genes include the heat shock promoters/enhancers from the *hsp70* and *hsp83* genes, useful for temperature induced expression; tissue specific promoters/enhancers such as the *sevenless* promoter/enhancer (Bowtell, et al., 1988, Genes Dev. 2(6):620-34), the *eyeless* promoter/enhancer (Bowtell, et al., 1991, Proc. Natl. Acad. Sci. U.S.A. 88(15):6853-7), and *glass*-responsive promoters/enhancers (Quiring, et al., 1994, Science 265:785-9) useful for expression in the eye; enhancers/promoters derived from the *dpp* or *vetigal* genes useful for expression in the wing (Staehling-Hampton, et al., 1994, Cell Growth Differ. 5(6):585-93; Kim, et al., 1996, Nature 382:133-8) and binary control systems employing exogenous DNA regulatory elements and exogenous transcriptional activator proteins, useful for testing the misexpression of genes in a wide variety of developmental stage-specific and tissue-specific patterns. Two examples of binary exogenous regulatory systems include the UAS/GAL4 system from yeast (Hay, et al., 1997, Proc. Natl. Acad. Sci. U.S.A. 94(10):5195-200; Ellis, et al., 1993, Development

119(3):855-65) and the "Tet system" derived from *E. coli*, which are described below. It is readily apparent to those skilled in the art that additional binary systems can be used which are based on other sets of exogenous transcriptional activators and cognate DNA regulatory elements in a manner similar to that for the UAS/GAL4 system and the Tet system.

In a specific embodiment, the UAS/GAL4 system is used. This system is a well-established and powerful method of mis-expression in *Drosophila* which employs the UAS_G upstream regulatory sequence for control of promoters by the yeast GAL4 transcriptional activator protein (Brand and Perrimon, 1993, Development 118(2):401-15). In this approach, transgenic *Drosophila*, termed "target" lines, are generated where the gene-of-interest (e.g. an insulin-like gene) to be mis-expressed is operably fused to an appropriate promoter controlled by UAS_G. Other transgenic *Drosophila* strains, termed "driver" lines, are generated where the GAL4 coding region is operably fused to promoters/enhancers that direct the expression of the GAL4 activator protein in specific tissues, such as the eye, wing, nervous system, gut, or musculature. The gene-of-interest is not expressed in the so-called target lines for lack of a transcriptional activator to "drive" transcription from the promoter joined to the gene-of-interest. However, when the UAS-target line is crossed with a GAL4 driver line, mis-expression of the gene-of-interest is induced in resulting progeny in a specific pattern that is characteristic for that GAL4 line. The technical simplicity of this approach makes it possible to sample the effects of directed mis-expression of the gene-of-interest in a wide variety of tissues by generating one transgenic target line with the gene-of-interest, and crossing that target line with a panel of pre-existing driver lines. A very large number of specific GAL4 driver lines have been generated previously and are available for use with this system.

In a second embodiment, a related method of directed mis-expression in *Drosophila* is used, that makes use of a tetracycline-regulated gene expression from *E. coli*, referred to as the "Tet system". In this case, transgenic *Drosophila* driver lines are generated where the coding region for a tetracycline-controlled transcriptional activator (tTA) is operably fused to promoters/enhancers that direct the expression of tTA in a tissue-specific and/or developmental stage-specific manner. Also, transgenic *Drosophila* target lines are generated where the coding region for the gene-of-interest to be mis-expressed (e.g. an insulin-like gene) is operably fused to a promoter that possesses a tTA-responsive regulatory element. Here again, mis-expression of the gene-of-interest can

be induced in progeny from a cross of the target line with any driver line of interest; moreover, the use of the Tet system as a binary control mechanism allows for an additional level of tight control in the resulting progeny of this cross. When *Drosophila* food is supplemented with a sufficient amount of tetracycline, it completely blocks expression of the gene-of-interest in the resulting progeny. Expression of the gene-of-interest can be induced at will simply by removal of tetracycline from the food. Also, the level of expression of the gene-of-interest can be adjusted by varying the level of tetracycline in the food. Thus, the use of the Tet system as a binary control mechanism for mis-expression has the advantage of providing a means to control the amplitude and timing of mis-expression of the gene-of-interest, in addition to spatial control. Consequently, if a gene-of-interest (e.g. an insulin-like gene) has lethal or deleterious effects when mis-expressed at an early stage in development, such as the embryonic or larval stages, the function of the gene-of-interest in the adult can still be assessed using the Tet system, by adding tetracycline to the food during early stages of development and removing tetracycline later so as to induce mis-expression only at the adult stage.

Analysis Of Mutant Phenotypes

After isolation of fruit flies carrying mutated or mis-expressed insulin-like genes, or inhibitory RNAs, animals are carefully examined for phenotypes-of-interest. For the situations involving deletions, insertions, point mutations, or mis-expression of insulin-like genes, fruit flies are generated that are homozygous and heterozygous for the altered insulin-like genes.

Examples of specific phenotypes that may be investigated include : altered body shape, altered body size, lethality, sterility, reduced brood size, increased brood size, altered life span, defective locomotion, alterted body plan, altered cell size, increased cell division, decreased cell division, altered feeding, slowed development, increased development, altered metabolism, (such as altered glycogen synthesis, storage, or degradation; altered lipid synthesis, storage or degradation; altered levels of carbohydrate in the hemolymph; and altered levels of lipid in the hemolymph), and altered morphogenesis of specific organs and tissues such as gonad, nervous system, fat body, hemocytes, peripheral sensory organs, bristles, imaginal discs, eye, wing, leg, antennae, gut, or musculature. For example, it is of particular interest to identify the ligand or ligands responsible for activating InR (or DIR), a *Drosophila* homologue of the insulin

receptor. A likely phenotype of a loss-of-function mutation in the ligand for the InR receptor might resemble one or more of the identified loss of function phenotypes for the receptor itself, including reduced body size and weight, reduced female fertility, increased developmental time, and/or defective embryonic neurogenesis.

Methods for creation and analysis of transgenic *Drosophila* strains having modified expression of genes are well known to those skilled in the art (Brand, et al., 1994, Methods in Cell Biology 44:635-654; Hay, et al., 1997, Proc. Natl. Acad. Sci. USA 94(10):5195-200). cDNAs or genomic regions encoding normal or mutant insulin-like genes can be operably fused to a desired promoter, as described above, and the promoter-insulin-like gene fusion inserted into any appropriate *Drosophila* transformation vector for the generation of transgenic flies. Typically, such transformation vectors are based on a well-characterized transposable elements, for example the P element (Rubin and Spradling, 1982, Science 218:348-53), the hobo element (Blackman, et al., 1989, Embo J. 8(1):211-7), mariner element (Lidholm, et al., 1993, Genetics 134(3):859-68), the hermes element (O'Brochta, et al., 1996, Genetics 142(3):907-14), Minos (Loukeris, et al., 1995, Proc. Natl. Acad. Sci. USA 92(21):9485-9), or the PiggyBac element (Handler, et al., 1998, Proc. Natl. Acad. Sci. USA 95(13):7520-5), where the terminal repeat sequences of the transposon that are required for transposition are incorporated into the transformation vector and arranged such that the terminal repeat sequences flank the transgene of interest (in this case a promoter-insulin-like gene fusion) as well as a marker gene used to identify transgenic animals. Most often, marker genes are used that affect the eye color of *Drosophila*, such as derivatives of the *Drosophila white* or *rosy* genes; however, in principle, any gene can be used as a marker that causes a reliable and easily scored phenotypic change in transgenic animals, and examples of other marker genes used for transformation include the Adh^{+} gene used as a selectable marker for the transformation of Adh^{-} strains, *Ddc*+ gene used to transform Ddc^{ts2} mutant strains, the *lacZ* gene of *E. coli*, and the $neomycin^{R}$ gene from the *E. coli* transposon Tn5. Plasmid constructs for introduction of the desired transgene are coinjected into *Drosophila* embryos having an appropriate genetic background, along with a helper plasmid that expresses the specific transposase need to mobilized the transgene into the genomic DNA. Animals arising from the injected embryos (G0 adults) are selected, or screened manually, for transgenic mosaic animals based on expression of the marker gene phenotype and are subsequently crossed to generate fully transgenic animals (G1 and subsequent generations)

that will stably carry one or more copies of the transgene of interest. Such stable transgenic animals are inspected for mutant phenotypes, such as abnormal development, morphology, metabolism, growth, longevity, reproduction, viability, or behavior, in order to determine a function for the insulin-like gene created by ectopic expression or overexpression of the insulin-like gene, or by expression of mutant insulin-like genes.

Generation of an overexpression/mis-expression phenotype is likely to result from either activation or inhibition of a receptor-linked signaling pathway. If such an overexpression/mis-expression phenotype is defined for an insulin-like gene, clonal analysis can then be used to determine whether this phenotype is restricted to cells expressing the insulin-like gene (i.e. whether the phenotype is cell autonomous or cell non-autonomous). Methods of mitotic recombination of chromosomes in heterozygous flies can be used to generate mitotic clones of genetically homozygous cells that are well known to those skilled in the art, which include the use of X-rays or preferably FLP/FRT mediated recombination (Xu and Harrison, 1994, Methods in Cell Biology 44:655-681; Greenspan, 1979, In *Fly Pushing: The Theory and Practice of Drosophila Genetics.* Plainview, NY, Cold Spring Harbor Laboratory Press: pp. 103-124). These mitotic recombination techniques result in patches of cells, mitotic clones, that contain 2 or no copies of the gene-of-interest. Production of the overexpression/mis-expression phenotype within cells in a clone having no copies of the gene-of-interest indicates that the effect is not cell autonomous, and is therefore likely to be the effect of a secreted molecule, as might be expected for insulin-like molecules.

Identification of Molecules that Interact With Insulin-Like Proteins

A variety of methods can be used to identify or screen for molecules, such as proteins or other molecules, that interact with insulin-like protein, or derivatives or fragments thereof. The assays may employ purified insulin-like protein, or cell lines or model organisms such as *Drosophila* and *C. elegans*, that have been genetically engineered to express insulin-like protein. Suitable screening methodologies are well known in the art to test for proteins and other molecules that interact with insulin-like gene and protein (see *e.g.*, PCT International Publication No. WO 96/34099). The newly identified interacting molecules may provide new targets for pharmaceutical or pesticidal agents. Any of a variety of exogenous molecules, both naturally occurring and/or synthetic (*e.g.*, libraries of small molecules or peptides, or phage display

libraries), may be screened for binding capacity. In a typical binding experiment, the insulin-like protein or fragment is mixed with candidate molecules under conditions conducive to binding, sufficient time is allowed for any binding to occur, and assays are performed to test for bound complexes. Assays to find interacting proteins can be performed by any method known in the art, for example, immunoprecipitation with an antibody that binds to the protein in a complex followed by analysis by size fractionation of the immunoprecipitated proteins (*e.g.* by denaturing or nondenaturing polyacrylamide gel electrophoresis), Western analysis, non-denaturing gel electrophoresis, *etc.*

Two-hybrid assay systems

A preferred method for identifying interacting proteins is a two-hybrid assay system or variation thereof (Fields and Song, Nature (1989) 340:245-246; U.S. Pat. No. 5,283,173; for review see Brent and Finley, Annu. Rev. Genet. (1997) 31:663-704). The most commonly used two-hybrid screen system is performed using yeast. All systems share three elements: 1) a gene that directs the synthesis of a "bait" protein fused to a DNA binding domain; 2) one or more "reporter" genes having an upstream binding site for the bait, and 3) a gene that directs the synthesis of a "prey" protein fused to an activation domain that activates transcription of the reporter gene. For the screening of proteins that interact with insulin-like protein, the "bait" is preferably a insulin-like protein, expressed as a fusion protein to a DNA binding domain; and the "prey" protein is a protein to be tested for ability to interact with the bait, and is expressed as a fusion protein to a transcription activation domain. The prey proteins can be obtained from recombinant biological libraries expressing random peptides.

The bait fusion protein can be constructed using any suitable DNA binding domain, such as the *E. coli* LexA repressor protein, or the yeast GAL4 protein (Bartel *et al.*, BioTechniques (1993) 14:920-924, Chasman *et al.*, Mol. Cell. Biol. (1989) 9:4746-4749; Ma *et al.*, Cell (1987) 48:847-853; Ptashne *et al.*, Nature (1990) 346:329-331).

The prey fusion protein can be constructed using any suitable activation domain such as GAL4, VP-16, *etc.* The preys may contain useful moieties such as nuclear localization signals (Ylikomi *et al.*, EMBO J. (1992) 11:3681-3694; Dingwall and Laskey, Trends Biochem. Sci. Trends Biochem. Sci. (1991) 16:479-481) or epitope tags

(Allen *et al.*, Trends Biochem. Sci. Trends Biochem. Sci. (1995) 20:511-516) to facilitate isolation of the encoded proteins.

Any reporter gene can be used that has a detectable phenotype such as reporter genes that allow cells expressing them to be selected by growth on appropriate medium (*e.g.* HIS3, LEU2 described by Chien *et al.*, PNAS (1991) 88:9572-9582; and Gyuris *et al.*, Cell (1993) 75:791-803). Other reporter genes, such as LacZ and GFP, allow cells expressing them to be visually screened (Chien *et al.*, *supra*).

Although the preferred host for two-hybrid screening is the yeast, the host cell in which the interaction assay and transcription of the reporter gene occurs can be any cell, such as mammalian (*e.g.* monkey, mouse, rat, human, bovine), chicken, bacterial, or insect cells. Various vectors and host strains for expression of the two fusion protein populations in yeast can be used (U.S. Pat. No. 5,468,614; Bartel *et al.*, Cellular Interactions in Development (1993) Hartley, ed., Practical Approach Series xviii, IRL Press at Oxford University Press, New York, NY, pp. 153-179; and Fields and Sternglanz, Trends In Genetics (1994) 10:286-292). As an example of a mammalian system, interaction of activation tagged VP16 derivatives with a GAL4-derived bait drives expression of reporters that direct the synthesis of hygromycin B phosphotransferase, chloramphenicol acetyltransferase, or CD4 cell surface antigen (Fearon *et al.*, PNAS (1992) 89:7958-7962). As another example, interaction of VP16-tagged derivatives with GAL4-derived baits drives the synthesis of SV40 T antigen, which in turn promotes the replication of the prey plasmid, which carries an SV40 origin (Vasavada *et al.*, PNAS (1991) 88:10686-10690).

Typically, the bait insulin-like gene and the prey library of chimeric genes are combined by mating the two yeast strains on solid or liquid media for a period of approximately 6-8 hours. The resulting diploids contain both kinds of chimeric genes, i.e., the DNA-binding domain fusion and the activation domain fusion.

Transcription of the reporter gene can be detected by a linked replication assay in the case of SV40 T antigen (described by Vasavada *et al.*, *supra*) or using immunoassay methods, preferably as described in Alam and Cook (Anal. Biochem. (1990)188:245-254). The activation of other reporter genes like URA3, HIS3, LYS2, or LEU2 enables the cells to grow in the absence of uracil, histidine, lysine, or leucine, respectively, and hence serves as a selectable marker. Other types of reporters are

monitored by measuring a detectable signal. For example, GFP and lacZ have gene products that are fluorescent and chromogenic, respectively.

After interacting proteins have been identified, the DNA sequences encoding the proteins can be isolated. In one method, the activation domain sequences or DNA-binding domain sequences (depending on the prey hybrid used) are amplified, for example, by PCR using pairs of oligonucleotide primers specific for the coding region of the DNA binding domain or activation domain. Other known amplification methods can be used, such as ligase chain reaction, use of Q replicase, or various other methods described (see Kricka *et al.*, Molecular Probing, Blotting, and Sequencing (1995) Academic Press, New York, Chapter 1 and Table IX). If a shuttle (yeast to *E. coli*) vector is used to express the fusion proteins, the DNA sequences encoding the proteins can be isolated by transformation of *E. coli* using the yeast DNA and recovering the plasmids from *E. coli*. Alternatively, the yeast vector can be isolated, and the insert encoding the fusion protein subcloned into a bacterial expression vector, for growth of the plasmid in *E. coli*.

Immunoassays

Immunoassays can be used to identify proteins that interact with or bind to insulin-like protein. Various assays are available for testing the ability of a protein to bind to or compete with binding to a wild-type insulin-like protein or for binding to an anti-insulin-like protein antibody. Suitable assays include radioimmunoassays, ELISA (enzyme linked immunosorbent assay), immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (*e.g.*, using colloidal gold, enzyme or radioisotope labels), western blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, immunoelectrophoresis assays, *etc.*

Biochemical Assays Using Insulin-Like Proteins

The present invention provides for biochemical assays using the insulin-like proteins. In one embodiment, *Drosophila* insulin-like proteins are useful for biochemical assays aimed at the identification and characterization of the ligand(s) for the known *Drosophila* insulin receptor encoded by the *InR* (DIR) gene (Nishida, et al., 1986,

Biochem. Biophys. Res. Commun. 141(2):474-81; Petruzzelli, et al., 1986, Proc. Natl. Acad. Sci. U.S.A. 83(13):4710-4; Fernandez-Almonacid and Rosen, 1987, Mol. Cell Biol. 7(8):2718-27), or the identification of ligands for new insulin-like receptor proteins that are discovered. The cDNAs encoding the insulin-like proteins can be individually subcloned into any of a large variety of eukaryotic expression vectors permitting expression in insect and mammalian cells, described above. The resulting genetically engineered cell lines expressing insulin-like proteins can be assayed for production, processing, and secretion of the mature insulin-like proteins, which lack the secretory signal peptide and connecting C peptide regions, for example with antibodies to *Drosophila* insulin-like proteins and Western blotting assays or ELISA assays. For assays of specific receptor binding and functional activation of receptor proteins, one can employ either crude culture medium or extracts containing secreted protein from genetically engineered cells (devoid of other insulin proteins), or partially purified culture medium or extracts, or preferably highly purified *Drosophila* insulin-like protein fractionated, for example, by chromatographic methods. Alternatively, mature *Drosophila* insulin-like protein can be synthesized using chemical methods (Nagata, et al., 1992, peptides 13(4):653-62).

Specific protein binding of *Drosophila* insulin-like proteins to the *Drosophila* InR receptor can be assayed as follows, for example, following the procedures of Yamaguchi et al. (Yamaguchi et al., 1995, Biochemistry 34:4962-4968). Chinese hamster ovary cells, COS cells, or any other suitable cell line, can be transiently transfected or stably transformed with expression constructs that direct the production of the *Drosophila* insulin receptor InR. Direct binding of a *Drosophila* insulin-like protein to such InR-expressing cells can be measured using a "labeled" purified *Drosophila* insulin-like protein derivative, where the label is typically a chemical or protein moiety covalently attached to the insulin-like polypeptide which permits the experimental monitoring and quantitation of the labeled *Drosophila* insulin-like protein in a complex mixture.

Specifically, the label attached to the insulin-like protein can be a radioactive substituent such as an ^{125}I-moiety or ^{32}P-phosphate moiety, a fluorescent chemical moiety, or labels which allow for indirect methods of detection such as a biotin-moiety for binding by avidin or streptavidin, an epitope-tag such as a Myc- or FLAG-tag, or a protein fusion domain which allows for direct or indirect enzymatic detection such as an alkaline phosphatase-fusion or Fc-fusion domain. Such labeled *Drosophila* insulin-like

proteins can be used to test for direct and specific binding to InR-expressing cells by incubating the labeled *Drosophila* insulin-like protein with the InR-expressing cells in serum-free medium, washing the cells with ice-cold phosphate buffered saline to remove unbound insulin-like protein, lysing the cells in buffer with an appropriate detergent, and measuring label in the lysates to determine the amount of bound insulin-like protein. Alternatively, in place of whole cells, membrane fractions obtained from InR-expressing cells may also be used. Also, instead of a direct binding assay, a competition binding assay may be used. For example, crude extracts or purified *Drosophila* insulin-like protein can be used as a competitor for binding of labeled purified bovine or porcine insulin to InR-expressing cells, by adding increasing concentrations of *Drosophila* insulin-like protein to the mixture. The specificity and affinity of binding of *Drosophila* insulin-like proteins can be judged by comparison with other insulin superfamily proteins tested in the same assay, for example vertebrate insulin, vertebrate IGF-I, vertebrate IGF-II, vertebrate relaxin, or silkmoth bombyxin.

Identification Of Additional Receptors Or Insulin-Like Binding Proteins

The invention described herein provides for methods in which Drosophila insulin-like proteins are used for the identification of novel insulin receptor proteins, other than *Drosophila* InR, using biochemical methods well known to those skilled in the art for detecting specific protein-protein interactions (Current Protocols in Protein Science, 1998, Coligan et al., eds., John Wiley & Sons, Inc., Somerset, New Jersey). Given the sequence diversity of the *Drosophila* insulin-like proteins detailed herein, the identification to date of only a single insulin receptor gene in *Drosophila, InR,* points to the possibility that some *Drosophila* insulin-like proteins may bind to other receptors. In particular, it is possible that some *Drosophila* insulin-like proteins interact with receptor types that have not yet been discovered in vertebrates, for example the relaxin receptor, or receptor types that are specific to invertebrates. The identification of either novel receptor types or invertebrate-specific receptor types is of great interest with respect to human therapeutic applications, or pesticide applications, respectively. Assuming some *Drosophila* insulin-like proteins do not exhibit specific protein binding to the known InR protein in the binding assays described above, then the novel cognate receptors for these insulin-like proteins can be investigated and identified as follows. Labeled *Drosophila* insulin-like proteins can be used for binding assays *in situ* to

identify tissues and cells possessing cognate receptors, for example as described elsewhere (Gorczyca et al., 1993, J. Neurosci. 13:3692-3704). Also, labeled *Drosophila* insulin-like proteins can be used to identify specific binding proteins including receptor proteins by affinity chromatography of *Drosophila* protein extracts using resins, beads, or chips with bound *Drosophila* insulin-like protein (Formosa, et al., 1991, Methods Enzymol 208:24-45; Formosa, et al., 1983, Proc. Natl. Acad. Sci. USA 80(9):2442-6). Further, specific insulin-binding proteins can be identified by cross-linking of radioactively-labeled or epitope-tagged insulin-like protein to specific binding proteins in lysates, followed by electrophoresis to identify and isolate the cross-linked protein species (Ransone, 1995, Methods Enzymol 254:491-7). Still further, molecular cloning methods can be used to identify novel receptors and binding proteins for *Drosophila* insulin-like proteins including expression cloning of specific receptors using *Drosophila* cDNA expression libraries transfected into mammalian cells, expression cloning of specific binding proteins using *Drosophila* cDNA libraries expressed in *E. coli* (Cheng and Flanagan, 1994, Cell 79(1):157-68), and yeast two-hybrid methods (as described above) using a *Drosophila* insulin-like protein fusion as a "bait" for screening activation-domain fusion libraries derived from *Drosophila* cDNA (Young and Davis, 1983, Science 222(4625):778-82; Young and Davis, 1983, Proc. Natl. Acad. Sci. USA 80(5):1194-8; Sikela and Hahn, 1987, Proc. Natl. Acad. Sci. USA 84(9):3038-42; Takemoto, et al., 1997, DNA Cell Biol 16(6):797-9).

ASSAYS OF INSULIN-LIKE PROTEINS

The functional activity of insulin-like proteins, derivatives and analogs can be assayed by various methods known to one skilled in the art.

For example, in one embodiment, where one is assaying for the ability to bind to or compete with a wild-type insulin-like protein for binding to an anti-insulin-like protein antibody, various immunoassays known in the art can be used, including competitive and non-competitive assay systems using techniques such as radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (*e.g.*, using colloidal gold, enzyme or radioisotope labels), western blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and

immunoelectrophoresis assays, *etc.* In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention. In another embodiment, where an insulin-like-binding protein is identified, the binding can be assayed, *e.g.*, by means well-known in the art. In another embodiment, physiological correlates of insulin-like protein binding to its substrates and/or receptors (*e.g.*, signal transduction) can be assayed.

In another embodiment, in insect (*e.g.*, Sf9 cells), fly (*e.g.*, *D. melanogaster*), or other model systems, genetic studies can be done to study the phenotypic effect of an insulin-like gene mutant that is a derivative or analog of a wild-type insulin-like gene. Other such methods will be readily apparent to the skilled artisan and are within the scope of the invention.

Other Functional Assays

For functional assays of *Drosophila* insulin-like protein, beyond receptor binding, the following activities can be investigated using InR-expressing cells after exposing said cells to crude or purified fractions of *Drosophila* insulin-like protein and comparing these results with those obtained with other insulin superfamily proteins described above (Yamaguchi et al., 1995, Biochemistry 34:4962-4968). Assayable functional activities include stimulation of cell proliferation; stimulation of overall tyrosine kinase activity by immunoblotting of cell extracts with an anti-phosphotyrosine antibody; stimulation of phosphorylation of specific substrate proteins such as InR or IRS-1 using ^{32}P- labeling and immunoprecipitation with antibodies that specifically recognize the substrate protein; and stimulation of other enzymatic activities linked to the insulin signaling pathway including assays of MAP kinase, Mek kinase, Akt kinase, and PI3-kinase activities.

Identifying Signaling Pathways And Phenotypes

This invention provides animal models which may be used in the identification and characterization of *D. melanogaster* insulin-like protein signaling pathways, and/or phenotypes associated with the mutation or abnormal expression of a *D. melanogaster* insulin-like protein. Methods of producing such animal models using novel genes and

proteins are well known in the art (*see e.g.*, PCT International Publication No. WO 96/34099, published October 31, 1996, which is incorporated by reference herein in its entirety). Such models include but are not limited to the following embodiments. Additional specific examples of animal models and their use are described below.

First, animals are provided in which a normal *D. melanogaster* insulin-like gene has been recombinantly introduced into the genome of the animal as an additional gene, under the regulation of either an exogenous or an endogenous promoter element, and as either a minigene or a large genomic fragment. Animals are also provided in which a normal gene has been recombinantly substituted for one or both copies of the animal's homologous gene by homologous recombination or gene targeting.

Second, animals are provided in which a mutant *D. melanogaster* insulin-like gene has been recombinantly introduced into the genome of the animal as an additional gene, under the regulation of either an exogenous or an endogenous promoter element, and as either a minigene or a large genomic fragment. Animals are also provided in which a mutant gene has been recombinantly substituted for one or both copies of the animal's homologous gene by homologous recombination or gene targeting.

Third, animals are provided in which a mutant version of one of that animal's own genes (bearing, for example, a specific mutation corresponding to, or similar to, a pathogenic mutation of an insulin-like gene from another species) has been recombinantly introduced into the genome of the animal as an additional gene, under the regulation of either an exogenous or an endogenous promoter element, and as either a minigene or a large genomic fragment.

Finally, equivalents of transgenic animals, including animals with mutated or inactivated genes, may be produced using chemical or x-ray mutagenesis. Using the isolated nucleic acids disclosed or otherwise enabled herein, one of ordinary skill may more rapidly screen the resulting offspring by, for example, direct sequencing, restriction fragment length polymorphism (RFLP) analysis, PCR, or hybridization analysis to detect mutants, or Southern blotting to demonstrate loss of one allele.

Such animal models may be used to identify a *D. melanogaster* insulin-like protein signaling pathway by various methods. In one embodiment, this invention provides a method of identifying a *D. melanogaster* insulin-like protein signaling pathway comprising: (a) disrupting a *D. melanogaster* insulin-like gene; and (b) identifying the effect of the gene disrupted in step (a) in an assay selected from the group consisting of a

developmental assay, an energy metabolism assay, a growth rate assay and a reproductive capacity assay, lethality, sterility, reduced brood size, increased brood size, altered life span, defective locomotion, altered body shape, altered body plan, altered body size, altered bristles, altered body weight, altered cell size, increased cell division, decreased cell division, altered feeding, slowed development, increased development, decreased metabolism (including alterations in glycogen synthesis, storage, and/or degradation, alterations in lipid synthesis, storage and/or degradation, alterations in levels of carbohydrate in hemolymph, alterations in levels of lipid in hemolymph), alterations in morphogenesis (including organs or tissues of the gonad, nervous system, fat body, hemacytes, peripheral sensory organs, imaginal discs, eye, wing, leg, antennae, bristle, gut or musculature). Such assays are well known to those skilled in the art. In one embodiment, results of the assay may be compared to known mutant phenotypes to determine the signaling pathway involved. In one embodiment, the gene is disrupted using chemical mutagenesis. In another embodiment, the gene is disrupted using transposon mutagenesis. In a further embodiment, the gene is disrupted by radiation mutagenesis.

Further, this invention provides a method of identifying a phenotype associated with mutation or abnormal expression of a *D. melanogaster* insulin-like protein comprising identifying the effect of a mutated or abnormally expressed *D. melanogaster* insulin-like gene in a *D. melanogaster* animal. In one embodiment, the effect is determined by any of the assays mentioned above in connection with identifying a *D. melanogaster* insulin-like protein signaling pathway. The gene may be mutated or abnormally expressed using any technique known in the art, such chemical mutagenesis, radiation mutagenesis, transposon mutagenesis, antisense and double-stranded RNA interference. Abnormal (i.e. ectopic) expression can be overexpression, underexpression (*e.g.*, due to inactivation), expression at a developmental time different from wild-type animals, or expression in a cell type different from in wild-type animals.

Analysis Of Genetic Interactions And Multiple Mutants

Yet another approach that may be used to probe the biological function of the insulin-like genes identified herein is by using tests for genetic interactions with other genes that may participate in the same, related, interacting, or modifying genetic or biochemical pathways. In particular, since it is evident that there are multiple insulin-like genes in the *Drosophila* genome, this raises the possibility of functional redundancy of one or more genes. Consequently, it is of interest to investigate the phenotypes of fruit flies

containing mutations that eliminate the function of more than one insulin-like gene. Such strains carrying mutations in multiple genes can be generated by cross breeding animals carrying the individual mutations, followed by selection of recombinant progeny that carry the desired multiple mutations.

One specific question-of-interest is genetic analysis of interactions of insulin-like genes with other well-characterized *Drosophila* genes and pathways. Thus, double mutant fruit flies may be constructed that carry mutations in an insulin-like gene and another gene-of-interest.

It is of particular interest to test the interaction of the insulin-like genes with other genes implicated in insulin signaling, especially those that exhibit homology to insulin signaling components in vertebrates. For example, fruit flies carrying mutations in insulin-like genes and either a loss-of-function mutation of *InR, chico, Pi3K92, Akt1, 14-3-3z, csw, Lar, Pk61C, Glut3, Ide, shaggy, s6k, Ras85D, drk, Sos, rl, or Dsor1* (FlyBase 1998, "FlyBase - A Drosophila Database", Nucleic Acids Research 26:85-88; http://flybase.bio.indiana.edu), would be of use in investigating the involvement of different insulin-like genes in the signaling pathway where these genes participate. Similarly, transgenic animals mis-expressing insulin-like genes which further carry mutations in the above-mentioned genes are of also of interest. Other genetic interactions may be tested based on the actual phenotypes observed for alterations of the insulin-like genes alone.

Genetic Modifier Screens

The initial characterization of phenotypes created by mutations in single or multiple insulin-like genes is expected to lead to the identification of *Drosophila* strains that exhibit mutant phenotypes suitable for large scale genetic modifier screens aimed at discovering other components of the same pathway. The procedures involved in typical genetic modifier screens to define other components of a genetic/biochemical pathway are well known to those skilled in the art and have been described elsewhere (Wolfner and Goldberg, 1994, Methods in Cell Biology 44:33-80; Karim et al., 1996, Genetics 143:315-329). Such genetic modifier screens are based on the identification of mutations in other genes that modify an initial mutant phenotype, by isolating either suppressor mutations that return the mutant phenotype toward normal, or enhancer mutations that make the initial mutant phenotype more severe.

Standard Genetic Modifier Screens

Genetic modifier screens differ depending upon the precise nature of the mutant allele being modified. If the mutant allele is genetically recessive, as is commonly the situation for a loss-of-function allele, then most typically males, or in some cases females, which carry one copy of the mutant allele are exposed to an effective mutagen, such as EMS, MMS, ENU, triethylamine, diepoxyalkanes, ICR-170, formaldehyde, X-rays, gamma rays, or ultraviolet radiation. The mutagenized animals are crossed to animals of the opposite sex that also carrying the mutant allele to be modified, and the resulting progeny are scored for rare events that result in a suppressed or enhanced version of the original mutant phenotype. In the case where the mutant allele being modified is genetically dominant, as is commonly the situation for ectopically expressed genes, wild type males are mutagenized and crossed to females carrying the mutant allele to be modified. Any new mutations identified as modifiers (i.e. suppressors or enhancers) are candidates for genes that participate in the same phenotype-generating pathway.

In a pilot-scale genetic modifier screen, 10,000 or fewer mutagenized progeny are inspected; in a moderate size screen, 10,000 to 50,000 mutagenized progeny are inspected; and in a large scale screen, over 50,000 mutagenized progeny are inspected. Progeny exhibiting either enhancement or suppression of the original phenotype are immediately crossed to adults containing balancer chromosomes and used as founders of a stable genetic line. In addition, progeny of the founder adult are retested under the original screening conditions to ensure stability and reproducibility of the phenotype. Additional secondary screens may be employed, as appropriate, to confirm the suitability of each new modifier mutant line for further analysis. For example, newly identified modifier mutations can be tested directly for interaction with other genes of interest known to be involved or implicated in insulin signaling pathways (*InR, chico, Pi3K92, Atk1, 14-3-3z, csw, Lar, Pk61C, Glut3, Ide, shaggy, s6k, Ras85D, drk, Sos, rl, Dsor1*, mutations in other insulin-like genes, or other modifier genes obtained from different genetic screens of the insulin signaling pathway), using methods described above. Also, the new modifier mutations can be tested for interactions with genes in other pathways thought to be unrelated or distantly related to insulin signaling, such as genes in the *Notch* signaling pathway. New modifier mutations that exhibit specific genetic interactions with other genes implicated in insulin signaling, but not interactions with genes in unrelated

pathways, are of particular interest. Additionally, strains can be generated that carry the new modifier mutations of interest in the absence of the original insulin-like gene mutation (i.e. a strain wild type for the mutant allele being suppressed or enhanced) to determine whether the new modifier mutation exhibits an intrinsic phenotype, independent of the mutation in the insulin-like gene, which might provide further clues as to the normal function of the new modifier gene.

Each newly-identified modifier mutation can be crossed to other modifier mutations identified in the same screen to place them into complementation groups, which typically correspond to individual genes (Greenspan, 1997, In *Fly Pushing: The Theory and Practice of Drosophila Genetics*, Plainview, NY, Cold Spring Harbor Laboratory Press: pp. 23-46). Two modifier mutations are said to fall within the same complementation group if animals carrying both mutations in *trans* exhibit essentially the same phenotype as animals that are homozygous for each mutation individually.

Gain-Of-Function Modifier Screens

Although the genetic modifier screens described above are quite powerful and sensitive, some genes that participate in an insulin-like pathway may be missed in this approach, particularly if there is functional redundancy of those genes. This is because the vast majority of the mutations generated in the standard mutagenesis methods described above will be loss-of-function mutations, whereas gain-of-function mutations that could reveal genes with functional redundancy will be relatively rare. Another method of genetic screening in *Drosophila* has been developed that focuses specifically on systematic gain-of- function genetic screens (Rorth, et al., 1998, Development 125:1049-1057). This method is based on a modular mis-expression system utilizing components of the GAL4/UAS system (which were defined above). In this case a modified P element, termed an EP element, is genetically engineered to contain a GAL4-responsive UAS element and promoter, and this engineered transposon is used to randomly tag genes by insertional mutagenesis (similar to the method of P mutagenesis described above). Thousands of transgenic *Drosophila* strains, termed EP lines, can thus be generated each containing a specific UAS-tagged gene. This approach takes advantage of a well-recognized insertional preference of P elements, where it has been found that P elements have a strong tendency to insert at the 5'-ends of genes. Consequently, many of the genes that have been tagged by insertion of EP elements become operably fused to a

GAL4-regulated promoter, and increased expression or mis-expression of the randomly tagged gene can be induced by crossing in a GAL4 driver gene (similar that described above).

Thus, systematic gain-of-function genetic screens for modifiers of phenotypes induced by mutation or mis-expression of an insulin-like gene can be performed as follows. A large battery of thousands of *Drosophila* EP lines can be crossed into a genetic background containing a mutant or mis-expressed insulin-like gene, and further containing an appropriate GAL4 driver transgene. The progeny of this cross can be inspected for enhancement or suppression of the original phenotype induced by mutation/mis-expression of the insulin-like gene. Progeny that exhibit an enhanced or suppressed phenotype can be crossed further to verify the reproducibility and specificity of this genetic interaction with the insulin-like gene. EP insertions that demonstrate a specific genetic interaction with a mutant or mis-expressed insulin-like gene, have therefore physically tagged a new gene that genetically interacts with the insulin-like. The new modifier gene can be identified and sequenced using PCR or hybridization screening methods that allow the isolation of the genomic DNA adjacent to the position of the EP element insertion.

Assays For Changes In Gene Expression

This invention provides assays for detecting changes in the expression of the *D. melanogaster* insulin-like genes and proteins. Assays for changes in gene expression are well known in the art (*see e.g.*, PCT Publication No. WO 96/34099, published October 31, 1996, which is incorporated by reference herein in its entirety). Such assays may be performed *in vitro* using transformed cell lines, immortalized cell lines, or recombinant cell lines, or *in vivo* using animal models.

In particular, the assays may detect the presence of increased or decreased expression of a *D. melanogaster* insulin-like gene or protein on the basis of increased or decreased mRNA expression (using, *e.g.*, nucleic acid probes), increased or decreased levels of related protein products (using, *e.g.*, the antibodies disclosed herein), or increased or decreased levels of expression of a marker gene (*e.g.*, -galactosidase or luciferase) operably linked to a 5' regulatory region in a recombinant construct.

In yet another series of embodiments, various expression analysis techniques may be used to identify genes which are differentially expressed between two conditions, such as a cell line or animal expressing a normal *D. melanogaster* insulin-like gene compared to

another cell line or animal expressing a mutant *D. melanogaster* insulin-like gene. Such techniques comprise any expression analysis technique known to one skilled in the art, including differential display, serial analysis of gene expression (SAGE), nucleic acid array technology, subtractive hybridization, proteome analysis and mass-spectrometry of two-dimensional protein gels. In a specific embodiment, nucleic acid array technology (*i.e.*, gene chips) may be used to determine a global (*i.e.*, genome-wide) gene expression pattern in a normal *D. melanogaster* animal for comparison with an animal having a mutation in one or more *D. melanogaster* insulin-like genes.

To elaborate further, the various methods of gene expression profiling mentioned above can be used to identify other genes (or proteins) that may have a functional relation to (*e.g.*, may participate in a signaling pathway with) a *D. melanogaster* insulin-like gene. Gene identification of such other genes is made by detecting changes in their expression levels following mutation, *i.e.*, insertion, deletion or substitution in, or overexpression, underexpression, mis-expression or knock-out, of a *D. melanogaster* insulin-like gene, as described herein. Expression profiling methods thus provide a powerful approach for analyzing the effects of mutation in a *D. melanogaster* insulin-like gene.

Insulin-Like Gene Regulatory Elements

This invention provides methods for using insulin-like gene regulatory DNA elements to identify tissues, cells, genes and factors that specifically control insulin-like protein production. In one embodiment, regulatory DNA elements, such as enhancers/promoters, from *Drosophila* insulin-like genes are useful for identifying and manipulating specific cells and tissues that synthesize an insulin-like protein. Such hormone secreting cells and tissues are of considerable interest since they are likely to have an important regulatory function within the animal in sensing and controlling growth, development, reproduction, and/or metabolism. Analyzing components that are specific to insulin-like protein secreting cells is likely to lead to an understanding of how to manipulate these regulatory processes, either for therapeutic applications or pesticide applications, as well as an understanding of how to diagnose dysfunction in these processes. For example, it is of specific interest to investigate whether there are neuroendocrine tissues in *Drosophila* that might have a function related to that of the mammalian pancreas in sensing and controlling metabolic activity through the production of an insulin-like protein. Regulatory DNA elements derived from insulin-like genes

provide a means to mark and manipulate such cells, and further, identify regulatory genes and proteins, as described below.

Gene Fusions With Insulin-Like Gene Regulatory DNA Elements

In a specific embodiment, gene fusions with the insulin-like regulatory elements can be made. For compact genes that have relatively few and small intervening sequences, such as the insulin-like genes described here, it is typically the case that the regulatory elements that control spatial and temporal expression patterns are found in the DNA immediately upstream of the coding region, extending to the nearest neighboring gene. Thus, putative regulatory DNA regions can be defined for the dIns2, dIns3, and dIns4 genes based on the sequence information provided in FIG.4. As shown in FIG. 4, the putative promoters ("PUT PROMOTER" or "PUT PROM") of the insulin-like genes are indicated with heavy lines below the respective sequences. Regulatory regions can be used to construct gene fusions where the regulatory DNAs are operably fused to a coding region for a reporter protein whose expression is easily detected, and these constructs are introduced as transgenes into *Drosophila*. An entire regulatory DNA region can be used, or the regulatory region can be divided into smaller segments to identify subelements that might be specific for controlling expression a given cell type or stage of development. Examples of reporter proteins that can be used for construction of these gene fusions include *E. coli* beta-galactosidase or the fluorescent GFP protein whose products can be detected readily in situ and which are useful for histological studies (O'Kane and Gehring, 1987, Proc. Natl. Acad. Sci. U.S.A. 84(24):9123-7; Chalfie, et al., 1994, Science 263:802-805) and sorting of specific cells that express insulin-like proteins (Cumberledge and Krasnow, 1994, Methods in Cell Biology 44:143-159); the cre or FLP recombinase proteins that can be used to control the presence and expression of other genes in the same cells through site-specific recombination (Golic and Lindquist, 1989, Cell 59(3):499-509; White, et al., 1996, Science 271:805-7); toxic proteins such as the reaper and hid cell death proteins which are useful to specifically ablate cells that normally express insulin-like proteins in order to assess the physiological function of this tissue (Kingston, 1998, In *Current Protocols in Molecular Biology*. Ausubel et al., John Wiley & Sons, Inc. sections 12.0.3-12.10) or any other protein where it is desired to examine the function this particular protein specifically in cells that synthesize and secrete insulin-like proteins (as described in the mis-expression analysis above).

Alternatively, a binary reporter system can be used, similar to that described above, where the insulin-like regulatory element is operably fused to the coding region of an exogenous transcriptional activator protein, such as the GAL4 or tTA activators described above, to create an insulin-like regulatory element "driver gene". For the other half of the binary system the exogenous activator controls a separate "target gene" containing a coding region of a reporter protein operably fused to a cognate regulatory element for the exogenous activator protein, such as UAS_G or a tTA-response element, respectively. An advantage of a binary system is that a single driver gene construct can be used to activate transcription from preconstructed target genes encoding different reporter proteins, each with its own uses as delineated above.

The insulin-like regulatory element-reporter gene fusions described in the preceding paragraph are also useful for tests of genetic interactions, where the objective is to identify those genes that have a specific role in controlling the expression of insulin-like genes, or promoting the growth and differentiation of the tissues that expresses the insulin-like protein. Transgenic *Drosophila* carrying an insulin-like regulatory element-reporter gene fusion can be crossed with another *Drosophila* strain carrying a mutation-of-interest and the resulting progeny examined. For example, the mutation-or-interest might be a modifier mutation arising from a genetic modifier screen as described in a preceding section. If no change of expression of the reporter gene in the resulting progeny is observed, this is indicative of a lack of involvement of the gene altered by the mutation-of-interest in controlling insulin-like protein expression; by contrast, if a significant increase, decrease, loss, or mis-expression of the reporter protein in the resulting progeny is observed, this is indicative of a regulatory role for the gene altered by the mutation-of-interest in cells expressing the insulin-like protein.

Protein-DNA Binding Assays

In a third embodiment, insulin-like gene regulatory DNA elements are also useful in protein-DNA binding assays to identify gene regulatory proteins that control the expression of insulin-like genes. Such gene regulatory proteins can be detected using a variety of methods that probe specific protein-DNA interactions well known to those skilled in the art (Kingston, 1998, In *Current Protocols in Molecular Biology*, Ausubel et al, John Wiley & Sons, Inc.. sections 12.0.3-12.10) including *in vivo* footprinting assays based on protection of DNA sequences from chemical and enzymatic modification within

living or permeabilized cells, *in vitro* footprinting assays based on protection of DNA sequences from chemical or enzymatic modification using protein extracts nitrocellulose filter-binding assays and gel electrophoresis mobility shift assays using radioactively labeled regulatory DNA elements mixed with protein extracts. In particular, it is of interest to identify those DNA binding proteins whose presence or absence is specific to insulin-like protein expressing tissue, as judged by comparison of the DNA-binding assays described above using cells/extracts from an insulin-like gene expressing tissue versus other cells/extracts from tissues that do not express insulin-like genes. For example, a DNA-binding activity that is specifically present in cells that normally express an insulin-like protein might function as a transcriptional activator of the insulin-like gene; conversely, a DNA-binding activity that is specifically absent in cells that normally express an insulin-like protein might function as a transcriptional repressor of the insulin-like gene. Having identified candidate insulin-like gene regulatory proteins using the above DNA-binding assays, these regulatory proteins can themselves by purified using a combination of conventional and DNA-affinity purification techniques. In this case, the DNA-affinity resins/beads are generated by covalent attachment to the resin of a small synthetic double stranded oligonucleotide corresponding to the recognition site of the DNA binding activity, or a small DNA fragment corresponding to the recognition site of the DNA binding activity, or a DNA segment containing tandemly iterated versions of the recognition site of the DNA binding activity. Alternatively, molecular cloning strategies can be used to identify proteins that specifically bind insulin-like gene regulatory DNA elements. For example, a *Drosophila* cDNA library in an *E. coli* expression vector, such as the lambda-gt11 vector, can be screened for *Drosophila* cDNAs that encode insulin-like gene regulatory element DNA-binding activity by probing the library with a labeled DNA fragment, or synthetic oligonucleotide, derived from the insulin-like gene regulatory DNA, preferably using a DNA region where specific protein binding has already been demonstrated with a protein-DNA binding assay described above (Singh et al., 1989, Biotechniques 7:252-61). Similarly, the yeast "one-hybrid" system can be used as another molecular cloning strategy (Li and Herskowitz, 1993, Science 262:1870-4; Luo, et al., 1996, Biotechniques 20(4):564-8; Vidal, et al., 1996, Proc. Natl. Acad. Sci. U.S.A. 93(19):10315-20). In this case, the insulin-like gene regulatory DNA element is operably fused as an upstream activating sequence (UAS) to one, or typically more, yeast reporter genes such as the *lacZ* gene, the *URA3* gene, the *LEU2* gene, the *HIS3* gene, or the *LYS2*

gene, and the reporter gene fusion construct(s) inserted into an appropriate yeast host strain. It is expected that in the engineered yeast host strain the reporter genes will not be transcriptionally active, for lack of a transcriptional activator protein to bind the UAS derived from the *Drosophila* insulin-like gene regulatory DNA. The engineered yeast host strain can be transformed with a library of *Drosophila* cDNAs inserted in a yeast activation domain fusion protein expression vector, e.g. pGAD, where the coding regions of the *Drosophila* cDNA inserts are fused to a functional yeast activation domain coding segment, such as those derived from the GAL4 or VP16 activators. Transformed yeast cells that acquire *Drosophila* cDNAs that encode proteins that bind the *Drosophila* insulin-like gene regulatory element can be identified based on the concerted activation the reporter genes, either by genetic selection for prototrophy (e.g. *LEU2*, *HIS3*, or *LYS2* reporters) or by screening with chromogenic substrates (*lacZ* reporter) by methods known in the art.

Use Of *Drosophila* Insulin-Like Proteins As A Media Supplement For Growth And Maintenance Of Insect Cells In Culture

Not all insect cells can be propagated effectively in available media and furthermore it is difficult and time consuming to wean cells onto serum-free media for large scale protein production. The present invention provides for the use of *Drosophila* homologs of insulin-like proteins, as media additive for growth and maintenance of cells in culture. Moreover, given that the *Drosophila* insulin-like proteins are the authentic endogenous protein hormones for *Drosophila* cells, and are likely to be more structurally and functionally similar to the authentic endogenous insulin-like hormones for other insect species, it is expected that *Drosophila* insulin-like hormones will exhibit superior properties in promoting growth and differentiation of insect cells in culture compared to the effects found for mammalian insulins on insect cells.

In a specific embodiment, the Drosophila insulin-like proteins are used for the in vitro cultivation of *Drosophila* or other insect cells. Insect cell lines are widely used for basic research on the cell and molecular biology of insects. Also, *Drosophila* and other insect cell lines have application as a preferred system for developing cell-based assays for insecticide targets, particularly those that might be amenable to high throughput screening methods (US Patent No. 5,767,261; US Patent No. 5,487,986; US Patent No. 5,641,652; US Patent No. 5,593,862; US Patent No. 5,593,864; US Patent No. 5,550,049; US Patent

No. 5,514,578).

In another embodiment, the *Drosophila* insulin-like proteins are employed for the *in vitro* cultivation of *Drosophila* and other insect cell lines used as host cells for the economical production of recombinant proteins on laboratory, pilot, or commercial scales. Further, the *Drosophila* and other insect cell lines can be used as hosts for the large-scale growth *in vitro* of viruses or bacteria that can be used as commercial insect control agents

Although fetal calf serum has been traditionally used as a media additive for the growth of insect cells in culture, it has a number of serious disadvantages. First, fetal calf serum is expensive, and is often used in large amounts at concentrations typically between 5% to 15%. Occasionally, fetal calf serum is not available commercially. Also, there are batch-to-batch variations in the activity of fetal calf serum in stimulating cell growth, and some batches have been found to be toxic to insect cells in culture. Thus, there is a need for substitutes for fetal calf serum in growth media for insect cells in culture, and the use of *Drosophila* insulin-like proteins for this purpose is expected to help fulfill this need.

Accordingly, *Drosophila* insulin-like proteins described herein can be used as an additive to insect cell growth media at concentrations preferably ranging from 5 ng/L to 0.5 g/L, and as a substitute for either fetal calf serum or mammalian insulin, for the following purposes (a) promoting the propagation of continuous insect cell lines from primary cultures; (b) promoting the differentiation and maintenance of specific insect cell types in culture such as nerve cells, muscle cells, or fat body cells; (c) promoting the propagation of insect cell lines *in vitro* for use in cell-based pesticide screening assays; (d) promoting the propagation of insect cell lines *in vitro* for use in large-scale production of recombinant proteins, natural protein products, or other natural products; and (e) promoting the propagation of insect cells for the large-scale production of viruses and bacteria which use insect cells as a host.

Agricultural Uses Of Drosophila Insulin-Like Genes

In another embodiment of the invention, *Drosophila* insulin-like genes may be used in controlling agriculturally important pest species. For example, the proteins disclosed herein, or analogs or derivatives thereof, may have activity in modifying the growth, feeding and/or reproduction of crop-damaging insects, or insect pests of farm animals or of other animals. In general, effective pesticides exert a disabling activity on the target pest such as lethality, sterility, paralysis, blocked development, or cessation of

feeding. Examples of such pests include egg, larval, juvenile and adult forms of flies, mosquitos, fleas, moths, beetles, cicadia, grasshoppers, and crickets.

Tests for such activities can be any method known in the art. Pesticides comprising the nucleic acids of the *Drosophila* insulin like proteins may be prepared in a suitable vector for delivery to a plant or animal. Examples of such vectors include *Agrobacterium tumefaciens* Ti plasmid-based vectors for the generation of transgenic plants or recombinant cauliflower mosaic virus for the incoulation of plant cells or retrovirus based vectors for the introduction of genes into vertebrate animals (Burns et al., 1993, Proc. Natl. Acad. Sci. USA 90:8033-37); and vectors based on transposable elements for the introduction of genes into insects. For example, transgenic insects can be generated using a transgene comprising an insulin-like gene operably fused to an appropriate inducible promoter. For example, a tTA-responsive promoter may be used in order to direct expression of the insulin-like protein at an appropriate time in the life cycle of the insect. In this way, one may test efficacy as an insecticide in, for example, the larval phase of the life cycle (i.e. when feeding does the greatest damage to crops).

Further, recombinant or synthetic insulin-like proteins, analogs, or derivatives can be assayed for insecticidal activity by injection of solutions of insulin-like proteins into the hemolymph of insect larvae (Blackburn, et al., 1998, Appl. Environ. Microbiol. 64(8):3036-41; Bowen and Ensign, 1998, Appl. Environ. Microbiol. 64(8):3029-35). Still further, transgenic plants that express insulin-like proteins can be tested for activity against insect pests (Estruch, et al., 1997, Nat. Biotechnol. 15(2):137-41).

In a preferred embodiment, insulin-like genes can be tested as insect control agents in the form of recombinant viruses that direct the expression of an insulin-like gene in the target pest. Suitable recombinant virus systems for expression of proteins in infected insect cells include recombinant Semliki Forest virus (DiCiommo and Bremner, 1998, J. Biol. Chem. 273:18060-66), recombinant sindbis virus (Higgs et al., 1995, Insect Mol. Biol. 4:97-103; Seabaugh et al., 1998, Virology 243:99-112), recombinant pantropic retrovirus (Matsubara et al., 1996, Proc. Natl. Acad. Sci. USA 93:6181-85; Jordan et al., 1998, Insect Mol. Biol. 7:215-22), and most preferably recombinant baculovirus. The use of recombinant baculovirus has a number of specific advantages including host specificity, environmental safety, the availability of easily manipulable vector systems, and the potential use of the recombinant virus directly as a pesticide without the need for purification or formulation of the insulin-like protein. Thus, recombinant baculoviruses

that direct the expression of insulin-like genes can be used for both testing the pesticidal activity of insulin-like proteins under controlled laboratory conditions, and as insect control agents in the field. One disadvantage of wild type baculoviruses as insect control agents can be the amount of time between application of the virus and death of the target insect, typically one to two weeks. During this period, the insect larvae continue to feed and damage crops. Consequently, there is a need to develop improved baculovirus-derived insect control agents which result in a rapid cessation of feeding of infected target insects. The well-known metabolic regulatory role of insulins in vertebrates raises the possibility that expression of insulin-like proteins from recombinant baculovirus in infected insects may have a desirable effect in controlling metabolism and limiting feeding of insect pests.

Mutational anaylsis of insulin-like genes may also be used in connection with the control of agriculturally-important pests. In this regard, mutational analysis of genes encoding insulin-like hormones in *Drosophila* provides a rational approach to determine the precise biological function of this class of hormones in invertebrates. Further, mutational analysis provides a means to validate potential pesticide targets that are constituents of these signaling pathways.

Drosophila insulin-like genes, proteins or derivatives thereof may be formulated with any carrier suitable for agricultural use, such as water, organic solvents and/or inorganic solvents. The pesticide composition may be in the form of a solid or liquid composition and may be prepared by fundamental formulation processes including dissolving, mixing, milling, granulating, and dispersing.

The present invention encompasses compositions containing a Drosophila insulin-like protein or gene in a mixture with agriculturally acceptable excipients known in the art, such as vehicles, carriers, binders, UV blockers, adhesives, hemecants, thickeners, dispersing agents, preservatives and insect attractants. Thus the compositions of the invention may, for example, be formulated as a solid comprising the active agent and a finely divided solid carrier. Alternatively, the active agent may be contained in liquid compositions including dispersions, emulsions and suspensions thereof. Any suitable final formulation may be used, including for example, granules, powder, bait pellets (a solid composition containing the active agent and an insect attractant or food substance), microcapsules, water dispersible granules, emulsions and emulsified concentrates.

Examples of adjuvant or carriers suitable for use with the present invention include

water, organic solvent, inorganic solvent, talc, pyrophyllite, synthetic fine silica, attapugus clay, kieselguhr chalk, diatomaceous earth, lime, calcium carbonate, bontonite, fuller's earth, cottonseed hulls, wheat flour, soybean flour, pumice, tripoli, wood flour, walnut shell flour, redwood flour, and lignin.

The compositions of the present invention may also include conventional insecticidal agents and/or may be applied in conjunction with conventional insecticidal agents.

EXAMPLES

Identification Of *D. Melanogaster* Insulin-Like Genes

A family of insulin-like genes has been identified in the model organism *D. melanogaster* (*i.e.*, the fly *Drosophila melanogaster*). This invention provides the following examples of identification of three *Drosophila* insulin-like genes as illustrated in the alignment of FIG. 8 and described in detail below.

Identification Of *Drosophila* Insulin-Like Genes In Genomic Sequence

A *Drosophila* cDNA encoding an insulin-like protein, termed dIns1, was identified by random sequencing of cDNAs in a library enriched for sequences expressed in the mesoderm (U.S. Patent Application Serial No. 09/201,226). We reasoned that other members of the insulin-like gene family in *Drosophila* could be identified by isolation and characterization of the genomic region surrounding the dINS1 gene.

Sequence database searches using the BLAST revealed that the dIns1 cDNA was identical over a 217 bp region to Dm3500, a sequence tagged site (STS) mapped by the Berkeley *Drosophila* Genome Project to chromosome 3, band 67C-D. Several P1 clones of genomic DNA had been molecularly mapped into a contig containing this STS, DS00060. Bacterial colonies containing P1 clones that molecularly map in and around DS00060 were obtained from Genome Systems, Inc. (St. Louis, Missouri). DNA from each of bacterial culture was screened for the presence of the dIns1 gene using a PCR-based assay. A small sample from each colony was picked with the end of a toothpick and transferred directly into 15 μl of PCR reaction buffer (supplied by the manufacturer, Perkin Elmer) containing 0.75 units Perkin Elmer Taq DNA polymerase, 2.5 mM $MgCl_2$, and 2.5 μM each of the following DNA primers:

LepEco5: CTA GGA ATT CGA TCG AGC AGG ATG AG (SEQ ID NO:8)
LepXba3: CAC TTC TAG ATC ATC AGG CGC AGT AG (SEQ ID NO:9)

Thermocycling conditions used were as follows (where 0:00 indicates time in minutes:seconds): an initial denaturation of 94°C, 4:00 followed by 35 cycles of 95°C, 0:30; 55°C, 1:00; and 72°C, 0:45. Products of the PCR reactions were analyzed by agarose gel electrophoresis. One of the P1 clones from this library, DS05250 (well L11, plate 14), was confirmed to produce a PCR product of the expected size for dINS1 and was selected for DNA sequencing.

The bacterial culture containing the DS05250 P1 clone was spread on an LB agar plate containing 25 μg/ml kanamycin, incubated overnight at 37°C, and a single colony was picked and used to inoculate 250 ml of Luria broth containing 25 μg/ml kanamycin. The culture was incubated with shaking at 37°C for 16 hours, bacterial cells were collected by centrifugation, and DNA was purified with a Qiagen Maxi-Prep System kit (QIAGEN, Inc., Valencia, California). The entire DNA sequence of the DS05250 P1 insert was obtained using a strategy that combined shotgun and directed sequencing of a small insert plasmid DNA library derived from the DS05250 P1 DNA (Ruddy DA, et al. Genome Research, 1997, 7:441-456). All DNA sequencing reactions were performed using standard protocols for the BigDye sequencing reagents (Applied Biosystems, Inc. Foster City, California) and products were analyzed using ABI 377 DNA sequencers. Trace data obtained from the ABI 377 DNA sequencers was analyzed and assembled into contigs comprising the complete P1 insert sequence using the phred-phrap computational package (Phil Green, U. of Washington).

Computational Strategy

The complete DNA sequence of the DS05250 P1 clone was analyzed by computational methods to identify insulin-like genes and other genes that might reside on this clone. The TBLASTN computer program (Altschul, et al., 1990, J. Mol. Biol. 215(3):403-10; Altschul, et al., 1997, Nucleic Acids Res. 25(17):3389-402) was employed with the dIns1 predicted protein sequence as a query to identify other insulin-like genes in this region. The results revealed that DS05250 contained part of the dIns1 coding region, as well as three other putative insulin-like genes in adjacent sequences (named dIns2,

dIns3, and dIns4; *see* FIG. 3). The GeneFinder (Phil Green, University of Washington) and GenScan programs (Burge and Karlin, 1997, J. Mol. Biol. 268(1):78-94) were used to predict coding regions, splice junctions, promoters, and poly(A) addition sites for each of the new insulin-like genes.

The presence of other gene sequences was investigated using the GeneFinder program, and also by analysis with the BLAST family of programs using the DS05250 sequence as a query against public and proprietary DNA and protein sequence databases. This analysis indicated that the DS05250 DNA contained additional genes distal to the dIns4 coding region with respect to the other insulin-like genes (FIG. 3); one region exhibited perfect homology to an uncharacterized *Drosophila* EST, and a second region exhibited a high degree of coding sequence homology with vertebrate anion channel proteins. Thus, we operationally defined the domain of the insulin-like multigene cluster in the DS05250 sequence as an 10,149 bp region that extends from the dIns1 end of the DNA insert to the start of the region homologous with the uncharacterized EST.

Since it was determined that the DS05250 P1 clone insert ended within the dIns1 gene and did not contain the complete cluster of insulin-like genes, a pooling strategy was employed using the remaining P1 clones mapped to this region in an effort to extend the sequence of the dIns1 end of this cluster. Accordingly, the following P1 clones were picked, pooled, and DNA prepared from bacterial cultures for DNA sequencing as described above for the DS05250 P1 clone: DS04166, DS07104, DS01000, DS06457, DS00683, DS00010, and DS00833. The same DNA sequencing strategy of combined shotgun and directed sequencing was employed on the pooled P1 clone DNA as that described above for the isolated DS0520 DNA. Individual sequence reads from the P1 pool were assembled with the DS05250 sequence contig using the phred-phrap computational package. The P1 pool strategy as successful in extending the sequence of the insulin-like gene cluster by 4.77 kbp beyond the end of the DS05250 sequence. Computational analysis of this additional sequence using the TBLASTN, GeneFinder, and GenScan programs, as above, revealed that the additional sequence from the P1 pool contained the N-terminal coding region of the dIns1 gene, an intergenic region, and an adjacent gene exhibiting homology to an uncharacteried *Drosophila* EST (*see* FIG. 3). Thus, we could define the limits of the cluster of repeated insulin-like genes in this genomic location as an 10,781 bp segment extending from the end of the sequences containing a predicted open reading frame with homology to the uncharacterized EST on

the dIns1 end of the cluster to the uncharacterized EST on the dIns3 end of the cluster (Fig. 4). An annotated sequence of the insulin multigene cluster in the DS05250 is presented in FIG. 4.

Isolation And Sequence Characterization Of Cdnas Corresponding To The *Drosophila* Insulin-Like Genes

The structure and expression of each new insulin-like gene predicted in the DS05250 genomic clone (dIns2, dIns3, and dIns4) was confirmed by either PCR amplification of inserts in *Drosophila* cDNA libraries, or reverse transcription of *Drosophila* mRNA and PCR amplification of the resulting cDNA (RT-PCR), as described below. For each gene, PCR primers were designed such that one primer annealed upstream of the predicted ATG codon, and the second primer annealed downstream of the predicted stop codon.

dIns2

The template source was a Canton S adult, oligo-dT- and random-primed cDNA library in the UniZap vector, purchased from Stratagene (Stratagene USA, LaJolla, California). Library DNA was diluted to a concentration of approximately 2 ng/μl before use. dIns2 cDNA was amplified by PCR, using a ClonTech Advantage cDNA PCR kit (CLONETECH Laboratories, Inc., Palo Alto, California) and the following primers:

fins2U70:	CTTCATCACTCATGGGCATCGAG	(SEQ ID NO:10)
fins2L515:	TGGGTTAATAGGTTTACGAGGTT	(SEQ ID NO:11)

The PCR reaction contained 1μl 10 X KlenTaq buffer, 1μl dNTPs, and 1μl KlenTaq enzyme mix, all as supplied by the manufacturer; to which was added 1 μl (2 ng) template DNA, and primers to a final concentration of 0.2 μM. Reaction conditions were as follows (where 0:00 indicates time in minutes:seconds): 95°C, 4:00, followed by 30 cycles of 95°C, 0:30; 55°C, 1:00; 68°C, 0:45.

Reaction products were analyzed by agarose gel electrophoresis, and a single major species was observed whose size matched that expected for the dIns2 cDNA (468 bp). The PCR product was isolated by electrophoresis in a 2% low melting point agarose gel stained with ethidium bromide, and the region of the gel containing the DNA was excised with a razor blade. Agarose was removed by digestion of the gel slice with ß-agarase as

follows: incubation at 65°C for 10 min, addition of approximately 1/10 vol. 10x ß-agarase buffer, brief incubation at 40°C, addition of 5 units ß-agarase, and incubation for 1 h at 40°C. The sample was quickly frozen in a dry ice/ethanol bath, and the remaining agarose removed by centrifugation in a microcentrifuge for 15 min. The supernatant was decanted and DNA precipitated by addition of sodium acetate to 0.3 M final concentration, a small amount of glycogen as carrier, and 2 volumes isopropanol. The mixture was left at -20°C for 30 min, and DNA collected by centrifugation in a microcentrifuge for 15 minutes. The resulting DNA pellet was dried and suspended in 10 μl TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA).

The purified dIns2 cDNA PCR product was cloned by ligation into the vector pCRII using the InVitrogen TA Cloning Kit (Invitrogen Corp., Carlsbad, California; Brun, et al., 1991, DNA Seq. 1(5):285-9) with subsequent transformation of *E. coli*, following the manufacturers directions. Individual transformant colonies were screened for the presence of the desired insert using a PCR assay with the dIns2-specific PCR primers (*i.e.* SEQ ID NO:10 and SEQ ID NO:11) described above. Plasmid DNA was isolated from the resulting colonies using an alkaline lysis method, and the insert DNA was sequenced using the BigDye sequencing kit (Applied Biosystems, Inc. Foster City, California) with universal M13 forward and reverse sequencing primers. The resulting sequence obtained for dIns2 cDNA (FIG. 5) was in agreement with that predicted from the DS05250 genomic sequence. Shown in Figure 5 is the annotated sequence of dIns2, which contains a signal sequence followed by a B peptide, C peptide, and A peptide, as indicated by the heavy lines below the respective sequences.

dIns3

The template source was freshly synthesized first strand cDNA generated using oligo-dT purified mRNA from 5 day old third instar larvae. cDNA synthesis was primed with oligo-dT primer containing a NotI site obtained from LifeTechnologies. The single stranded cDNA was amplified by PCR, using the ClonTech Advantage cDNA PCR kit and the following primers designed from the predicted dIns3 genomic sequence:

fins3OU16:	GCTTCCGATTTAGTGGTATAAA	(SEQ ID NO:12)
fins3OL584:	TTCGTATGTATGTATGTATGTG	(SEQ ID NO:13)

The PCR reaction contained 1 μl 10 X KlenTaq buffer, 1 μl dNTPs, and 1μl KlenTaq enzyme mix, all as supplied by the manufacturer; to which was added 0.5 μl template DNA and primers at a final concentration of 0.2 μM. Reaction conditions were as follows (where 0:00 indicates time in minutes:seconds): 95°C, 4:00, followed by 30 cycles of 95°C, 0:30; 55°C, 1:00; 68°C, 0:45.

The reaction products were analyzed by gel electrophoresis and a single major species of the size expected for dIns3 was observed. The dIns3 cDNA product was cloned into the vector pCRII as described above for dIns2.

The dIns3 cDNA inserts in pCRII clones were sequenced by PCR amplification of the insert DNA with either M13 forward and reverse primers, or fins3OU16 and fins3OL584 primers, followed by cycle-sequencing of the amplification products. The sequence determined for the dIns3 cDNA clones (FIG. 6) was in agreement with that predicted from the genomic sequence derived from the DS05250 P1 clone. Shown in Figure 6 is the annotated sequence of dIns3, which contains a signal sequence followed by a B peptide, C peptide, and A peptide, as indicated by the heavy lines below the respective sequences.

dIns4

Reverse transcription and PCR amplification were used to obtain dIns4 cDNA clones as described above for dIns3 except that the following primers were used:

fins4U5:	TAAACCCATAACCATGAGCAAGC	(SEQ ID NO:14)
fins4L516:	TCAGTTGGGGTCAATGATTTTCG	(SEQ ID NO:15)

A single major product of the expected size was observed following agarose gel electrophoresis and the resulting dIns4 cDNA was purified, cloned and sequenced as described above for dIns2. The sequence determined for the dIns4 cDNA clone (FIG. 7) was in agreement with that predicted from the genomic sequence derived from the DS05250 P1 clone. Shown in Figure 7 is the annotated sequence of dIns4, which contains a signal sequence followed by a B peptide, C peptide, and A peptide, as indicated by the heavy lines below the respective sequences.

Structural Features Of Drosophila Insulin-Like Genes And Proteins

The genomic organization of *Drosophila* insulin-like genes revealed in the DS05250 sequence can be viewed as two pairs of genes, dIns1/dIns2 and dIns3/dIns4, where the genes in each pair are arranged in tandem and oriented in the same direction, but where each pair of genes is oriented in the opposite direction and transcribed convergently (*see* FIG. 3). This implies that during the evolution of this multigene cluster an inversion occurred to create this arrangement, as opposed to the simplest model for the generation of a multigene array resulting solely from unequal cross-over, which would produce tandem genes all oriented in the same direction (Kondo, et al., 1996, J. Mol. Biol. 259:926-937; Smit, et al., 1998, Prog. Neurobiol. 54:35-54).

The sequence of the genomic region of DS05250 also reveals that three of the four *Drosophila* insulin-like genes, dIns1, dIns2, and dIns4, have intervening sequences that disrupt coding regions. It is notable that the position of the intervening sequence is at essentially the same location in each of these genes: within the C peptide coding sequences very near the junction with the B peptide coding sequences (FIG. 4). This same approximate position of an intervening sequence is also frequently found in vertebrate insulin-like genes, supporting an evolutionary relationship between *Drosophila* and vertebrate members of the insulin superfamily (Murray-Rust, et al., 1992, BioEssays 14:325-331; McRory and Sherwood, 1997, DNA and Cell Biology 16:939-949). The dIns3 gene does not appear to have an intervening sequence that disrupts the coding region of this gene. There is precedent for this situation in the form of the bombyxin genes of *Lepidoptera*, which all lack intervening sequences (Kondo, et al., 1996, J. Mol. Biol. 259:926-937).

Alignment of the predicted sequences of the *Drosophila* insulin-like proteins with other vertebrate and invertebrate members of the insulin superfamily demonstrates that the *Drosophila* proteins all contain the key structural features known to be important for promoting proper folding and processing of these preprohormones (FIG. 8). It is particularly notable that each of the *Drosophila* insulin-like proteins (dIns1, dIns2, dIns3 and dIns4) possesses a large C peptide of more than 30 amino acids flanked by dibasic residues, which are recognized by prohormone convertases during removal of the C peptide from the prohormone. Also, none of the *Drosophila* insulin-like proteins have a large C- terminal extension, such as found in the E peptide region of IGFs. Consequently, the overall organization of the *Drosophila* insulin-like proteins is similar to that of vertebrate insulins rather than that of vertebrate IGFs, although the possibility remains that

one or more *Drosophila* insulin-like proteins might have a growth-promoting function similar to that of vertebrate IGFs. This is of interest since it remains uncertain when the structure and function of IGFs diverged from insulins during metazoan evolution (McRory and Sherwood, 1997, DNA and Cell Biology 16:939-949). Also, the *Drosophila* insulin-like receptor InR exhibits a ligand-specificity with a preference for insulins as opposed to IGFs, even though InR appears to mediate growth-promoting activities in vivo.

All of the *Drosophila* insulin-like proteins possess exactly the same number (six) and spacing of Cys residues as found in vertebrate insulin superfamily proteins (boxed in FIG. 8), indicating that the disulfide bonding pattern stabilizing the folded structure of these proteins would also be identical. This contrasts with the situation for some other invertebrate insulin-like proteins which have been found to have unusual disulfide features including an extra pair of Cys residues (represented in FIG. 8 by MIP-I from freshwater snail, and F13B12 from the nematode *C. elegans*) or which may lack the conserved Cys residues (Brousseau, et al., 1998, Early 1998 East Coast Worm Meeting, abstract 20; Duret, et al., 1998, Genome Res. 8(4):348-53; Wisotzkey and Liu, 1998, Early 1998 East Coast Worm Meeting, abstract 206; Pierce and Ruvkun, 1998, Early 1998 East Coast Worm Meeting, abstract 150), or have altered spacing between Cys residues in the A or B chains (found in some *C. elegans* insulin-like proteins, (Kondo, et al., 1996, J. Mol. Biol. 259:926- 937; Smit, et al., 1998, Prog. Neurobiol. 54:35-54). It is also evident that all of the *Drosophila* insulin-like proteins have hydrophobic residues in positions that normally contribute to stabilizing the core structure at the interface between the A and B peptides in the folded protein (shaded in FIG. 8). Given the presence of these conserved structural features in each of the *Drosophila* insulin-like proteins it is expected that they will adopt a secondary and tertiary structure very similar to that found in their vertebrate and invertebrate counterparts, specifically a long central helix in the B peptide and two short antiparallel helices in the A peptide joined by a loop.

Despite the presence of such conserved structural features, phylogenetic analyses indicate that the *Drosophila* insulin-like proteins are rather diverse at the primary sequence level, particularly at positions expected to be exposed on the surface of the mature hormones. This is all the more surprising given that these *Drosophila* genes are located immediately next to one another in the genome, and might therefore be expected to have evolved relatively recently from each other. By contrast, the very large family of known bombyxin proteins in *Lepidoptera* exhibits considerably less sequence divergence than the

family of four *Drosophila* insulin-like proteins discussed here. Similarly, the family of five insulin-like proteins found in the freshwater snail, the MIPS, are also less diverse at the protein sequence level than the four *Drosophila* insulin-like proteins. Indeed, the *Drosophila* insulin-like proteins are more divergent from each other than the degree of sequence divergence observed between vertebrate insulins and IGFs. Accordingly, this sequence divergence amoung the *Drosophila* insulin-like proteins suggests the possibility that they may serve distinctly different functions and/or act by binding through different receptor proteins.

Cross-Hybridization Experiment For dIns1, dIns2, dIns3, and dIns4

Sequence alignments of the four *Drosophila* insulin-like proteins revealed diversity among these family members at the amino acid level (*see* Fig. 8). Computational comparisons of the nucleic acid sequences using BLASTN and dot plot programs provided further evidence of sequence divergence in both coding and non-coding regions. As an experimental demonstration of the sequence divergence of the dIns1 genes, a Southern blotting experiment was performed where the dIns1 cDNA was used as probe to test cross-hybridization with the other *Drosophila* insulin-like genes, and a *C. elegans* insulin-like gene (F13B12), under conditions of low, medium, and high stringency, as described below.

Plasmid DNAs (0.5 g) containing inserts of each insulin-like cDNA were digested with an appropriate restriction enzyme to liberate the cDNA insert from the vector as follows: pcDNA3.1-dInS1, PmeI; pBS+-dIns2, EcoRI, pBS+-dIns3, EcoRI, pBS+-dIns4, EcoRI; and pcDNA3.1-F13B12, PmeI. The restriction enzyme digestions were divided into thirds (for testing low, medium and high stringency hybridization), arranged in three identical sets, and the products were separated by electrophoresis in a 1% agarose gel along with DNA size markers. DNA fragments were visualized by staining with ethidium bromide, UV transillumination, and photography. The results demonstrated complete digestion of each plasmid DNA, and importantly showed that approximately the same amount of each insulin-like cDNA fragment was present in the gel. DNAs in the gel were denatured by treatment with a 0.4 N NaOH solution, blotted to a Hybond N+ membrane (Amersham) by transfer in the same solution, and the membrane neutralized by washing in a buffer containing 0.5 M Tris-HCl, pH 7.2, 1 M NaCl. The membrane was cut into thirds, each containing an identical set of the different insulin-like cDNAS, and the

membranes were pretreated in a hybridization buffer (0.5 M sodium phosphate. pH 7.2, 7% SDS, 1 mM EDTA, and 1% bovine serum albumin) which also contained 100 g/ml sheared, denatured salmon sperm DNA. A DNA probe was prepared by digestion of a plasmid vector containing dIns1 cDNA with EcoRI to release the insert, separation of the dIns1 cDNA from the vector by agarose gel electrophoresis, and radiolabelling using -^{32}P-dCTP with an Amersham Rediprime DNA labelling kit following the manufacturers directions (final incorporation of radioactivity into the probe was 30 Ci). Hybridization of the probe to membranes was carried out by incubating each membrane in the hybridization buffer above along with 10 Ci of ^{32}P-labeled dIns1 cDNA probe overnight at 45°C. After hybridization, each membrane was washed two times each for 30 minutes each at 45°C in wash buffer #1 (40 mM sodium phosphate, pH 7,2, 5% SDS, 1 mM EDTA, 0.5% bovine serum albumin), followed by four washes each for 30 minutes in wash buffer #2 (40 mM sodium phosphate, pH 7.2, 1% SDS, 1 mM EDTA), and subsequently each membrane was treated differently as described below for low, medium, or high stringency hybridization conditions. For low stringency hybridization, one membrane was not washed further. For medium stringency hybridization, a second membrane was subjected to four washes each for 30 minutes in wash buffer #2 at 55°C. For high stringency hybridization, a third membrane was subjected to four washes each for 30 minutes in wash buffer #2 at 55°C, followed by four washes each for 30 minutes in wash buffer #2 at 65°C. The membranes were dried and radioactivity detected by autoradiography using X-ray film and an intensifying screen. Hybridization of the ^{32}P-labeled dIns1 cDNA probe to the homologous dIns1 cDNA on the membranes was readily detected after only 15 minutes of autoradiography under all three hybridization conditions, and increasing the time of autoradiography to 2.5 hours revealed no detectable cross-hybridation of dIns1 probe to the dIns2, dIns3, dIns4, or F13B12 cDNAs on the membranes under any hybridization condition. After 2.5 hours of autoradiography, very weak hybridization of the probe could be detected to pBS+ vector fragments and marker DNA fragments, which was most evident on the low and medium stringency membranes (presumably due to weak nonspecific hybridization). Thus, these results clearly demonstrate that there is no significant cross-hybridization of dIns1 cDNA to any of the other *Drosophila* insulin-like cDNAs, dIns2, dIns3, and dIns4, under conditions of either low, medium or high stringency. Furthermore, these results provide clear experimental evidence of the significant sequence divergence of these genes.

WHAT IS CLAIMED IS:

1. A purified protein comprising an amino acid sequence of an A peptide domain and/or B peptide domain of a *Drosophila* insulin-like protein selected from dIns2, dIns3 and dIns4, as depicted in Figure 8.

2. The purified protein of claim 1 comprising an amino acid sequence selected from any one of SEQ ID Nos 2, 4, and 6 and 8.

3. A purified antibody, or derivative thereof containing the idiotype of the antibody, capable of immunospecific binding to the protein of Claim 1 and not to an insulin-like protein of another species.

4. An isolated nucleic acid comprising a nucleotide sequence encoding an amino acid sequence as depicted in Figure 5 (SEQ ID NO:2), Figure 6 (SEQ ID NO:4), or Figure 7 (SEQ ID NO:6), wherein said nucleic acid is less than 15 kilobases.

5. The isolated nucleic acid of Claim 4 comprising a nucleotide sequence selected from SEQ ID NO: 1, 3, and 5.

6. A recombinant cell containing a recombinant nucleic acid vector comprising a nucleotide sequence encoding an amino acid sequence as depicted in Figure 5 (SEQ ID NO:2), Figure 6 (SEQ ID NO:4), or Figure 7 (SEQ ID NO:6).

7. A method of identifying a phenotype associated with mutation or abnormal expression of a *D. melanogaster* insulin-like protein comprising identifying an effect of a mutated or abnormally expressed *D. melanogaster* insulin-like gene which encodes a *D. melanogaster* insulin-like protein comprising an amino acid sequence as depicted in Figure 5 (SEQ ID NO:2), Figure 6 (SEQ ID NO:4), or Figure 7 (SEQ ID NO:6), in a *D. melanogaster* animal.

8. The method of Claim 7 wherein the gene is mutated or abnormally expressed using a technique selected from the group consisting of radiation mutagenesis, chemical mutagenesis, transposon mutagenesis, antisense and double-stranded RNA interference.

9. A modified, isolated *D. melanogaster* animal in which a *D. melanogaster* insulin-like gene which encodes a *D. melanogaster* insulin-like protein comprising an amino acid sequence as depicted in Figure 5 (SEQ ID NO:2), Figure 6 (SEQ ID NO:4), or Figure 7 (SEQ ID NO:6) which has been deleted or inactivated by recombinant methods, or a progeny thereof containing the deleted or inactivated gene.

10. A recombinant non-human animal containing a *D. melanogaster* insulin-like transgene which encodes a *D. melanogaster* insulin-like protein comprising an amino acid sequence as depicted in Figure 5 (SEQ ID NO:2), Figure 6 (SEQ ID NO:4), or Figure 7 (SEQ ID NO:6).

11. A method of identifying a molecule that alters the expression level of a *D. melanogaster* insulin-like gene corresponding to a cDNA sequence as depicted in Figure 5 (SEQ ID NO:1), Figure 6 (SEQ ID NO:3), or Figure 7 (SEQ ID NO:5), which method comprises:

(a) contacting a transgenic fly cell with one or more molecules, said transgenic fly cell having a transgene comprising a promoter or enhancer region of genomic DNA from 1 base to 6 kilobases upstream of the start codon of the cDNA sequence operably linked to a reporter gene; and

(b) determining whether the level of expression of the reporter gene is altered relative to the level of expression of the reporter gene in the absence of the one or more molecules.

12. A cell culture medium or medium supplement comprising (a) a sterile liquid carrier, and (b) a protein or fragment thereof, functional in promoting cell growth, survival, or differentiation, said protein comprising at least 10 contiguous amino acids as depicted in Figure 5 (SEQ ID NO:2), Figure 6 (SEQ ID NO:4), or Figure 7 (SEQ ID NO:6).

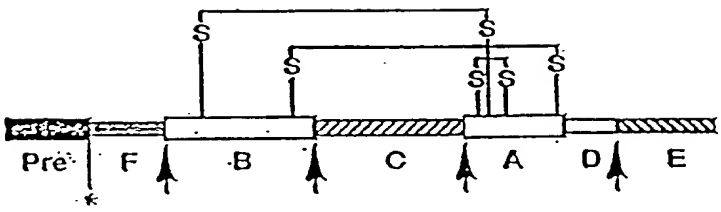

FIG. 1

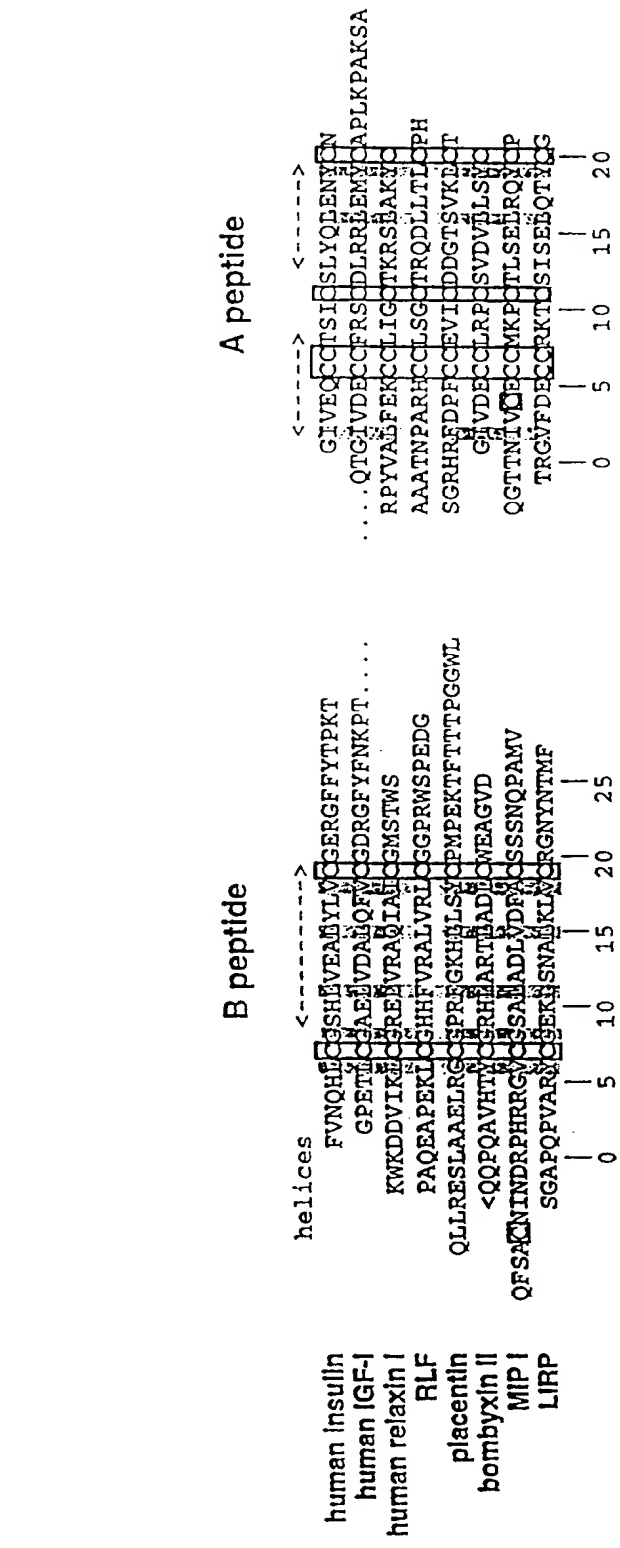

FIG. 2

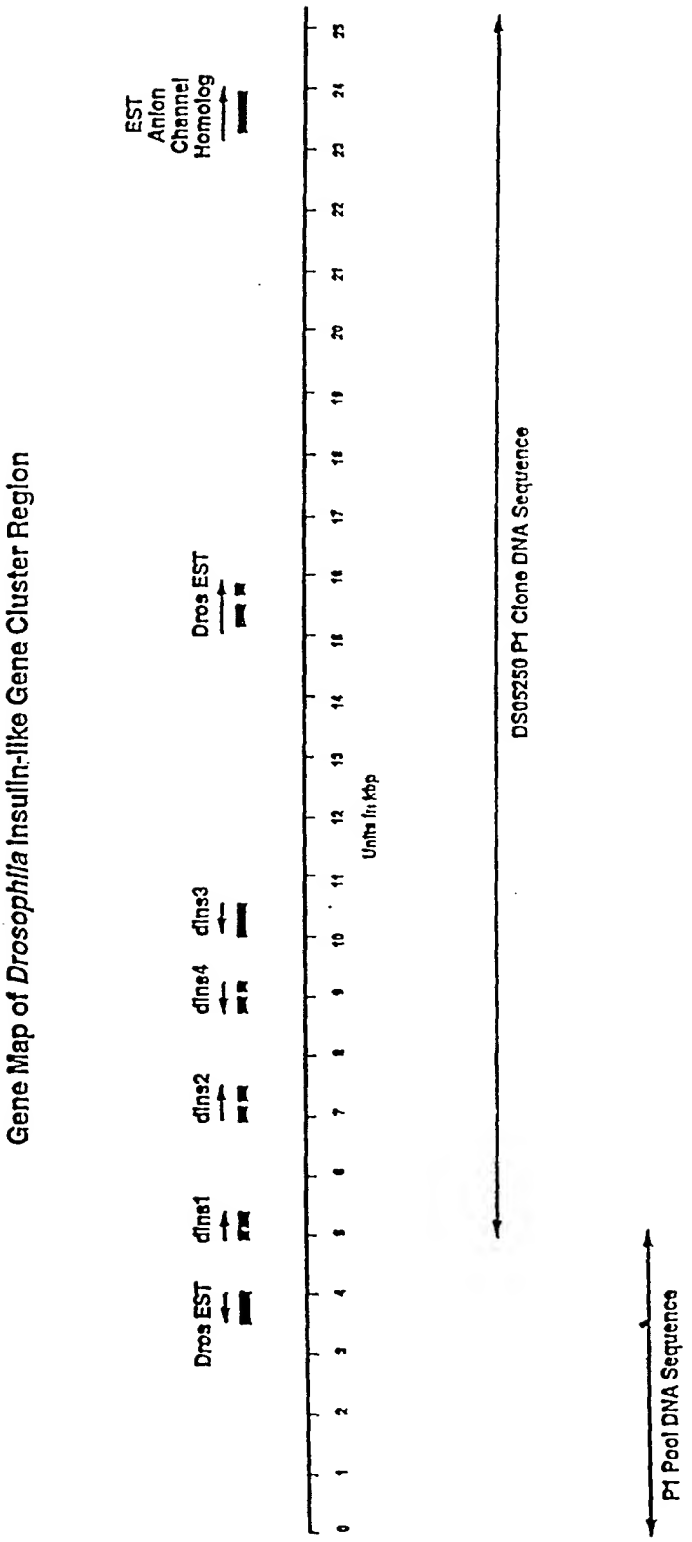

FIG. 3

Drosophila Insulin-like Gene Cluster Genomic DNA Sequenc

```
        10         20         30         40         50         60
GCTTCTGCTCGGAGAGCGGCTGACCCGAATGGGATAGGGCATCTCCTGTCCACAGGAACG
CGAAGACGAGCCTCTCGCCGACTGGGCTTACCCTATCCCGTAGAGGACAGGTGTCCTTGC

        70         80         90        100        110        120
AACTCCCCATTATCGCCCTGCTTGGCCTTCATCGTCTTCACCGCCGATTCAATCAGTCTC
TTGAGGGGTAATAGCGGGACGAACCGGAAGTAGCAGAAGTGGCGGCTAAGTTAGTCAGAG

       130        140        150        160        170        180
CAAGGACTATCGCAGTTGTAGTGTCCAAAGGGACCAGGTGGCTCCGATGCCGCACTATTG
GTTCCTGATAGCGTCAACATCACAGGTTTCCCTGGTCCACCGAGGCTACGGCGTGATAAC

       190        200        210        220        230        240
GCATTGGAGCCGGAAACCGATCGTGCTAACTCCCAGCAGCTTCTGTATATGTCGCCCTGC
CGTAACCTCGGCCTTTGGCTAGCACGATTGAGGGTCGTCGAAGACATATACAGCGGGACG

       250        260        270        280        290        300
GTGGAGTAGAGTGTGTCAAGGTGGAGGTCACTAATGTGCCAGAAGTAGCCTGTCAACAAA
CACCTCATCTCACACAGTTCCACCTCCAGTGATTACACGGTCTTCATCGGACAGTTGTTT

       310        320        330        340        350        360
CAAGTAGAATCAAGTAAATGTGTTAGTTAAATACCCATAGATATATGTAAAAGTTGTTGT
GTTCATCTTAGTTCATTTACACAATCAATTTATGGGTATCTATATACATTTTCAACAACA

       370        380        390        400        410        420
TTTATTTGCTAAGAAAAGTTTAATCTATATCCCAGTTTTACACACCAGATTTTTATGTCC
AAATAAACGATTCTTTTCAAATTAGATATAGGGTCAAAATGTGTGGTCTAAAAATACAGG

       430        440        450        460        470        480
TGAGCAATTTCGTATGTATTTCCCCTTCGTAAAGTAAGGATCGAGATTAGACTTTGACTT
ACTCGTTAAAGCATACATAAAGGGGAAGCATTTCATTCCTAGCTCTAATCTGAAACTGAA

       490        500        510        520        530        540
TGGTTAAGTCGGGCAATTCCTGGCCGGGAAAGGCCATTTCCTTTCGCGGGGCATTTTCCC
ACCAATTCAGCCCGTTAAGGACCGGCCCTTTCCGGTAAAGGAAAGCGCCCCGTAAAAGGG

       550        560        570        580        590        600
GCCGGCTGGTCGAGCGACAAAAATAAGAAAAACCTGGTAGTTCAAATGGAAATCTCCTGC
CGGCCGACCAGCTCGCTGTTTTTATTCTTTTTGGACCATCAAGTTTACCTTTAGAGGACG

       610        620        630        640        650        660
AGCTGACTGTTTGGTTGGTTGACTGACCTGGCCCGAATTTAACTTTCTACCTGGTCGCAA
TCGACTGACAAACCAACCAACTGACTGGACCGGGCTTAAATTGAAAGATGGACCAGCGTT

       670        680        690        700        710        720
TACGTGAAGTCAAAAAGTCAATTAGCGAGTCAACATTTTGAGCGCCGGCCAACTCCAAGG
ATGCACTTCAGTTTTTCAGTTAATCGCTCAGTTGTAAAACTCGCGGCCGGTTGAGGTTCC
                                    _____DINS1 PUT.  PROMOTER_______>
```

FIG. 4A

```
        730         740         750         760         770         780
ATCAGTATCATTTGGCATGCCCAGCGATCGGTTTGCCAAGAGCACGAGAAGTTCGAGATA
TAGTCATAGTAAACCGTACGGGTCGCTAGCCAAACGGTTCTCGTGCTCTTCAAGCTCTAT
___________>

        790         800         810         820         830         840
GGACCCAGAGATACCAGAGATAAAGGAGGCATACCTTTTATGCCCGGTGAGAGCACGGAC
CCTGGGTCTCTATGGTCTCTATTTCCTCCGTATGGAAAATACGGGCCACTCTCGTGCCTG

        850         860         870         880         890         900
GGCGGAGTGAAAGATCGAGCAGGATGAGCCTGATTAGACTGGGAGTGGCGCTGCTGCTCC
CCGCCTCACTTTCTAGCTCGTCCTACTCGGACTAATCTGACCCTGACCGCGACGACGAGG
                          M   S   L   I   R   L   G   L   A   L   L   L>
                        ____DINS1 CODING REGION (EXON 1)_____>

        910         920         930         940         950         960
TGCTGGCCACCGTGTCGCAGTTACTGCAGCCGGTCCAGGGACGCCGAAAGATGTGCGGCG
ACGACCGGTGGCACAGCGTCAATGACGTCGGCCAGGTCCCTGCGGCTTTCTACACGCCGC
L   L   A   T   V   S   Q   L   L   Q   P   V   Q   G   R   R   K   M   C   G>
_________________DINS1 CODING REGION (EXON 1)_________________>

>End_of_DS05250_sequence
        |
        970         980         990        1000        1010        1020
AGGCTCTGATCCAGGCACTGGATGTGATTTGTGTTAATGGATTTACACGCCGTGTCAGGC
TCCGAGACTAGGTCCGTGACCTACACTAAACACAATTACCTAAATGTGCGGCACAGTCCG
E   A   L   I   Q   A   L   D   V   I   C   V   N   G   F   T   R   R   V   R>
_________________DINS1 CODING REGION (EXON 1)_________________>

       1030        1040        1050        1060        1070        1080
GGAGCAGTGGTAAGTTTGGGTACTATGCATATTCGATTGGCTTCCATACATCTAACTTCT
CCTCGTCACCATTCAAACCCATGATACGTATAAGCTAACCGAAGGTATGTAGATTGAAGA
R   S   S   A>
__________>

       1090        1100        1110        1120        1130        1140
TTTCGACAAGCGTCTAAGGATGCTAGAGTGCGAGACCTTATCCGTAAGCTACAGCAGCCG
AAAGCTGTTCGCAGATTCCTACGATCTCACGCTCTGGAATAGGCATTCGATGTCGTCGGC
                S   K   D   A   R   V   R   D   L   I   R   K   L   Q   Q   P>
             ___________DINS1 CODING REGION (EXON-2)_____________>

       1150        1160        1170        1180        1190        1200
GATGAGGACATTGAACAGGAAACGGAAACGGGAAGGTTAAAGCAGAAGCATACGGATGCG
CTACTCCTGTAACTTGTCCTTTGCCTTTGCCCTTCCAATTTCGTCTTCGTATGCCTACGC
 D   E   D   I   E   Q   E   T   E   T   G   R   L   K   Q   K   H   T   D   A>
_________________DINS1 CODING REGION (EXON-2)_________________>

       1210        1220        1230        1240        1250        1260
GATACGGAGAAGGGTGTGCCACCGGCCGTCGGAAGTGGACGAAAGTTGCGACGCCATCGG
CTATGCCTCTTCCCACACGGTGGCCGGCAGCCTTCACCTGCTTTCAACGCTGCGGTAGCC
 D   T   E   K   G   V   P   P   A   V   G   S   G   R   K   L   R   R   H   R>
_________________DINS1 CODING REGION (EXON-2)_________________>
```

FIG. 4B

```
      1270       1280       1290       1300       1310       1320
CGACGCATCGCCCACGAGTGTTGCAAGGAGGGCTGCACCTACGACGATATACTGGACTAC
GCTGCGTAGCGGGTGCTCACAACGTTCCTCCCGACGTGGATGCTGCTATATGACCTGATG
 R  R  I  A  H  E  C  C  K  E  G  C  T  Y  D  D  I  L  D  Y>
_________________DINS1 CODING REGION (EXON-2)_________________>

                          >dIns1_poly(A)_signal
                                          |
      1330       1340       1350       | 1360       1370       1380
TGCGCCTGATGACCAGGATGGCAAAACAAAACAAATAAAAACCAGAAACCAGATCCCAAA
ACGCGGACTACTGGTCCTACCGTTTTGTTTTGTTTATTTTTGGTCTTTGGTCTAGGGTTT
 C  A  *>
__________>

      1390       1400       1410       1420       1430       1440
AACCAAGTACCAGATGAACACGACATGGCTGAGATTTTGTGTGGCGGCACGGGGAAAACA
TTGGTTCATGGTCTACTTGTGCTGTACCGACTCTAAAACACACCGCCGTGCCCCTTTTGT

      1450       1460       1470       1480       1490       1500
CCCGACGACCGGCAGGCTATTTGCAATTCATTTTCCTACTACACTTAACCCCTAACTATA
GGGCTGCTGGCCGTCCGATAAACGTTAAGTAAAAGGATGATGTGAATTGGGGATTGATAT

      1510       1520       1530       1540       1550       1560
AACGTAATCGTATTTCCAAATATTTCATTGTAAAATTTCTAGTGGAGGCAAATAAAGTTA
TTGCATTAGCATAAAGGTTTATAAAGTAACATTTTAAAGATCACCTCCGTTTATTTCAAT

      1570       1580       1590       1600       1610       1620
CTCTCCAAGCAGCAGCAGAAACAAAAGAAGAGTCCATTGCTTTTTTCTACATTCTACGCC
GAGAGGTTCGTCGTCGTCTTTGTTTTCTTCTCAGGTAACGAAAAAAGATGTAAGATGCGG

      1630       1640       1650       1660       1670       1680
CTGCAGCATTCCAGCTGTGAGGCATGGGGAATCCCCTTGTTATTCAAACCACCCGAAGCC
GACGTCGTAAGGTCGACACTCCGTACCCCTTAGGGGAACAATAAGTTTGGTGGGCTTCGG

      1690       1700       1710       1720       1730       1740
ACCCAAACCATCGAGCCACCCACAAGCAGCTGCCATTCAGCACCTCGAGTGCGGTGCCCT
TGGGTTTGGTAGCTCGGTGGGTGTTCGTCGACGGTAAGTCGTGGAGCTCACGCCACGGGA

      1750       1760       1770       1780       1790       1800
TGTTTTCCGAGAACAATAATGAAAAATATGAATTTTTAATTAGATGACGTTCTGATTTTA
ACAAAAGGCTCTTGTTATTACTTTTTATACTTAAAAATTAATCTACTGCAAGACTAAAAT

      1810       1820       1830       1840       1850       1860
ATAAGCAAAACAAAAGGTGGAGACAAAACGAACTCGGTAATACACTCAGATTCGAATTTA
TATTCGTTTTGTTTTCCACCTCTGTTTTGCTTGAGCCATTATGTGAGTCTAAGCTTAAAT

      1870       1880       1890       1900       1910       1920
CAGCTTCCTTTTTATCCATAATTTTTGTTATTATCGAAGGAGCGATATCAAAACTAGAAA
GTCGAAGGAAAAATAGGTATTAAAAACAATAATAGCTTCCTCGCTATAGTTTTGATCTTT

      1930       1940       1950       1960       1970       1980
ACAACTTCCAATCAGTAGCGGGATTTTCCGAAGATAACACTCTATTCAACCGAAGGGTTT
TGTTGAAGGTTAGTCATCGCCCTAAAAGGCTTCTATTGTGAGATAAGTTGGCTTCCCAAA
```

FIG. 4C

```
      1990        2000        2010        2020        2030        2040
TGAAATGATAATAATTCCGTTCTTACAGGTAAAAATCTATACTAATACCTGTTTTTTTGC
ACTTTACTATTATTAAGGCAAGAATGTCCATTTTTAGATATGATTATGGACAAAAAAACG

      2050        2060        2070        2080        2090        2100
GGACGGAAAAAAGGCTCAGTTGGCTTATCATTGGCAAAAGGGACTTGGGGAAACCATAAA
CCTGCCTTTTTTCCGAGTCAACCGAATAGTAACCGTTTTCCCTGAACCCCTTTGGTATTT

      2110        2120        2130        2140        2150        2160
GTATCGAAGGTACTGAGCCAAGATAATGAGATAACAGAAGGCGACTTTATTGTTTTCCAC
CATAGCTTCCATGACTCGGTTCTATTACTCTATTGTCTTCCGCTGAAATAACAAAAGGTG

      2170        2180        2190        2200        2210        2220
TCAAAAGCAATTGAATAAGTTGGCACTCGTTTTTAATTGAATGGGAATGAAATAAGCTCT
AGTTTTCGTTAACTTATTCAACCGTGAGCAAAAATTAACTTACCCTTACTTTATTCGAGA

      2230        2240        2250        2260        2270        2280
AAAAGTGTTGTTAAAACGTAATGGCTTTTGTGTTAATTTAAAGAATTTAAGTAGTTTTGA
TTTTCACAACAATTTTGCATTACCGAAAACACAATTAAATTTCTTAAATTCATCAAAACT

      2290        2300        2310        2320        2330        2340
AAGTATCATTATTCTTTAGGTAATTTTTATTACATTCCAAATTTAATAAATGACTAATTC
TTCATAGTAATAAGAAATCCATTAAAAATAATGTAAGGTTTAAATTATTTACTGATTAAG

      2350        2360        2370        2380        2390        2400
GAAAAAGTGTTTATTTAATCAATGAATATATTTCAAGTAAGTTTACTTTTAGTAGCTTGC
CTTTTTCACAAATAAATTAGTTACTTATATAAAGTTCATTCAAATGAAAATCATCGAACG

      2410        2420        2430        2440        2450        2460
CAAATGTGAGTTTAAATATGTATGCATAGAACTATATAGTTAAACTGCTAAACTTTACAG
GTTTACACTCAAATTTATACATACGTATCTTGATATATCAATTTGACGATTTGAAATGTC

      2470        2480        2490        2500        2510        2520
TTAAACTTTCTGAACCCACCAAAATGGATGAACATCCTCGTCTGCCGAAGGGAACTCGAT
AATTTGAAAGACTTGGGTGGTTTTACCTACTTGTAGGAGCAGACGGCTTCCCTTGAGCTA

      2530        2540        2550        2560        2570        2580
GCACGTCATTTTGTTTTTCAACAATCCAGATCCGTGCGCTACTCCTTGGGCGAGAAAGTA
CGTGCAGTAAAACAAAAAGTTGTTAGGTCTAGGCACGCGATGAGGAACCCGCTCTTTCAT

      2590        2600        2610        2620        2630        2640
AACAAACGCCAGCTGATATGCGTCAGACCCCCCGGGCTCATCATCATCTCACCATTTCAG
TTGTTTGCGGTCGACTATACGCAGTCTGGGGGGCCCGAGTAGTAGTAGAGTGGTAAAGTC

      2650        2660        2670        2680        2690        2700
ACATCCCATGCCAGCCCGAATCCTCACGAGAAACTAGACCAGACCAGGGCGAACTACATA
TGTAGGGTACGGTCGGGCTTAGGAGTGCTCTTTGATCTGGTCTGGTCCCGCTTGATGTAT

      2710        2720        2730        2740        2750        2760
TGTGGATGATGCTAACTGACACTACGGCTGACTCATGCTGACAGTGCTCAGACGCTGGAT
ACACCTACTACGATTGACTGTGATGCCGACTGAGTACGACTGTCACGAGTCTGCGACCTA
```

FIG. 4D

```
      2770       2780       2790       2800       2810       2820
ACAGCCCGCAGACATCCAACTCGTATCCTATCCGATTCTGCCCCATATATATAACCCTCA
TGTCGGGCGTCTGTAGGTTGAGCATAGGATAGGCTAAGACGGGGTATATATATTGGGAGT
                                                 ____DINS2 PUT. _____>

      2830       2840       2850       2860       2870       2880
GTCGATGGCTGGGAGGCAAACAGTTGAGGCCGTGCCACTTGGCAGACACATACTACACAC
CAGCTACCGACCCTCCGTTTGTCAACTCCGGCACGGTGAACCGTCTGTGTATGATGTGTG
____DINS2 PUT. PROM_____>

      2890       2900       2910       2920       2930       2940
TCCCCGGGGGATTCACGCATCCATACTTAAACACCACTTCATCACTCATGGGCATCGAGA
AGGGGCCCCCTAAGTGCGTAGGTATGAATTTGTGGTGAAGTAGTGAGTACCCGTAGCTCT
                                                             _>

      2950       2960       2970       2980       2990       3000
TGAGGTGTCAGGACAGGAGGATCCTGCTACCTAGCCTACTCCTACTAATCCTTATGATCG
ACTCCACAGTCCTGTCCTCCTAGGACGATGGATCGGATGAGGATGATTAGGAATACTAGC
M  R  C  Q  D  R  R  I  L  L  P  S  L  L  L  L  I  L  M  I>
_________________DINS2 CODING REGION (EXON-1)_____________________>

      3010       3020       3030       3040       3050       3060
GCGGTGTCCAGGCCACCATGAAGTTGTGCGGCCGCAAACTGCCCGAAACTCTCTCCAAGC
CGCCACAGGTCCGGTGGTACTTCAACACGCCGGCGTTTGACGGGCTTTGAGAGAGGTTCG
G  G  V  Q  A  T  M  K  L  C  G  R  K  L  P  E  T  L  S  K>
_________________DINS2 CODING REGION (EXON-1)_____________________>

      3070       3080       3090       3100       3110       3120
TCTGTGTGTATGGCTTCAACGCAATGACCAAGAGAACTTTGGGTAGGTGGGATTTTTCTT
AGACACACATACCGAAGTTGCGTTACTGGTTCTCTTGAAACCCATCCACCCTAAAAAGAA
L  C  V  Y  G  F  N  A  M  T  K  R  T  L  D>
_______DINS2 CODING REGION (EXON-1)_________>

      3130       3140       3150       3160       3170       3180
GATATAAGGAATACTAAAGTGCCATATCTCTTTACTTTCACCTAACACCTGTAGACCCCG
CTATATTCCTTATGATTTCACGGTATAGAGAAATGAAAGTGGATTGTGGACATCTGGGGC
                                                           P>
                                                          _____>

      3190       3200       3210       3220       3230       3240
TGAACTTCAATCAGATCGATGGCTTCGAAGACCGTTCCCTGCTGGAAAGACTGTTGAGTG
ACTTGAAGTTAGTCTAGCTACCGAAGCTTCTGGCAAGGGACGACCTTTCTGACAACTCAC
V  N  F  N  Q  I  D  G  F  E  D  R  S  L  L  E  R  L  L  S>
_________________DINS2 CODING REGION (EXON-2)_____________________>

      3250       3260       3270       3280       3290       3300
ATAGTTCGGTTCAGATGCTCAAGACTCGACGTCTTCGGGATGGAGTCTTCGACGAGTGTT
TATCAAGCCAAGTCTACGAGTTCTGAGCTGCAGAAGCCCTACCTCAGAAGCTGCTCACAA
D  S  S  V  Q  M  L  K  T  R  R  L  R  D  G  V  F  D  E  C>
_________________DINS2 CODING REGION (EXON-2)_____________________>
```

FIG. 4E

```
      3310       3320       3330       3340       3350       3360
GCCTGAAGTCGTGCACCATGGATGAGGTGCTGAGATATTGTGCTGCCAAGCCGAGAACGT
CGGACTTCAGCACGTGGTACCTACTCCACGACTCTATAACACGACGGTTCGGCTCTTGCA
C  L  K  S  C  T  M  D  E  V  L  R  Y  C  A  A  K  P  R  T>
_________________DINS2 CODING REGION (EXON-2)___________________>

      3370       3380       3390       3400       3410       3420
AAACCTCGTAAACCTATTAACCCAATGACGACAACTGCGATGATTGAAATGGAATGAAAG
TTTGGAGCATTTGGATAATTGGGTTACTGCTGTTGACGCTACTAACTTTACCTTACTTTC
*>
__>

      3430       3440       3450       3460       3470       3480
GACCCGATTGGGGAAAGCACTCACGTAATCATAGTTGTTAAGTCGTTATCGAAGCCTACT
CTGGGCTAACCCCTTTCGTGAGTGCATTAGTATCAACAATTCAGCAATAGCTTCGGATGA

      3490       3500       3510       3520       3530       3540
CAATTCCAACTTTGGATTTATGATATATATGCACATGTAAGAGGGATGTATGCGCATAAT
GTTAAGGTTGAAACCTAAATACTATATATACGTGTACATTCTCCCTACATACGCGTATTA

      3550       3560       3570       3580       3590       3600
TTATGATCTGAAATCAGAGACAGGCACGCGAAATGAATCGGAACACGGGATGTTATGCAT
AATACTAGACTTTAGTCTCTGTCCGTGCGCTTTACTTAGCCTTGTGCCCTACAATACGTA

                                          >dIns2_put._poly(A)_signal
                                                              |
      3610       3620       3630       3640       3650       3660
GGTAGATATGTATGATTGTGCGGGGCCAGAATACATCGCCTGGGTATAAATTATTAAATA
CCATCTATACATACTAACACGCCCCGGTCTTATGTAGCGGACCCATATTTAATAATTTAT

      3670       3680       3690       3700       3710       3720
AATTATGTATTCAAACTGCTGCAGATTGGCCAACTTGATTGGTAATGAAACGGGTATTAC
TTAATACATAAGTTTGACGACGTCTAACCGGTTGAACTAACCATTACTTTGCCCATAATG

                                  <dIns4_put._poly(A)_signal
                                                  |
      3730       3740       3750       3760|      3770       3780
ATTGATTTTTCATTGTCGTTCATTGCAGTTAATTATTTATTGAACAGCGGCCGGATTTCT
TAACTAAAAAGTAACAGCAAGTAACGTCAATTAATAAATAACTTGTCGCCGGCCTAAAGA

      3790       3800       3810       3820       3830       3840
GTTTGCAACTATGTTGAAAAGGAAGCTGTGATTTTTTAACAAACTCTGTTCATTGTAAAG
CAAACGTTGATACAACTTTTCCTTCGACACTAAAAAATTGTTTGAGACAAGTAACATTTC

      3850       3860       3870       3880       3890       3900
TTTAAAATCATTCCAATTTAATGCCCTCAAAACCTACGCTGAAATGGTCAGTTTTAAAAC
AAATTTTAGTAAGGTTAAATTACGGGAGTTTTGGATGCGACTTTACCAGTCAAAATTTTG

      3910       3920       3930       3940       3950       3960
GATATTTATTAATATTTTAGTTAATTTACTAAGATTATCCGTTTTGCACTTTTAATGCCT
CTATAAATAATTATAAAATCAATTAAATGATTCTAATAGGCAAAACGTGAAAATTACGGA
```

FIG. 4F

```
      3970       3980       3990       4000       4010       4020
TGCATTTGGTAATGCGTGATTGTTATTTAAGGTCTGCATGAATTTAGTTGATTCCGTTTA
ACGTAAACCATTACGCACTAACAATAAATTCCAGACGTACTTAAATCAACTAAGGCAAAT

      4030       4040       4050       4060       4070       4080
TTTTAGCTTTCAAAATGTAATAATCTTCTAATTTACAACTACACAGAACGATTAAATTAT
AAAATCGAAAGTTTTACATTATTAGAAGATTAAATGTTGATGTGTCTTGCTAATTTAATA

      4090       4100       4110       4120       4130       4140
GAGTATTGCTATAAAATCGGCCAACCGCGACTAGAAATACTCGACTTTTAAGGTCAACAT
CTCATAACGATATTTTAGCCGGTTGGCGCTGATCTTTATGAGCTGAAAATTCCAGTTGTA

      4150       4160       4170       4180       4190       4200
AAAAGTAAGTCAATGTTTTGATTATAAGATTTGATCAATTACTTCTTTACGGATGATATA
TTTTCATTCAGTTACAAAACTAATATTCTAAACTAGTTAATGAAGAAATGCCTACTATAT

      4210       4220       4230       4240       4250       4260
ATCATCGATAAACGAAGTACGAAAAAAGCTATGAACTAAAATTTGGAAATTTCCCACATG
TAGTAGCTATTTGCTTCATGCTTTTTTCGATACTTGATTTTAAACCTTTAAAGGGTGTAC

      4270       4280       4290       4300       4310       4320
CGACTAACTTTTGAATTGCAATTGGATTGCCTACTGTATTAAGACAGAAACAAGTTTTGG
GCTGATTGAAAACTTAACGTTAACCTAACGGATGACATAATTCTGTCTTTGTTCAAAACC

      4330       4340       4350       4360       4370       4380
AAATGAATGAATGGTTTAAATTGTTTCAAGTTTTTTTAAGATTTTTTTTGTTTTCAATAA
TTTACTTACTTACCAAATTTAACAAAGTTCAAAAAAATTCTAAAAAAAACAAAAGTTATT

      4390       4400       4410       4420       4430       4440
ATTTAGTTTTAATAGAAAAAAAGATATATTCATTTTAGATTTCTGAATACTTGTGTTATA
TAAATCAAAATTATCTTTTTTTCTATATAAGTAAAATCTAAAGACTTATGAACACAATAT

      4450       4460       4470       4480       4490       4500
TCGCTTTTTATTCAAGTGTAATAATCAACATATATATCATATAATGATAATAATAAATGT
AGCGAAAAATAAGTTCACATTATTAGTTGTATATATAGTATATTACTATTATTATTTACA

      4510       4520       4530       4540       4550       4560
AACGTCCCAAATTAATAATAATATAAAGTAGCATTTGCGATTGTTTGCCAAAGCTTAAAG
TTGCAGGGTTTAATTATTATTATATTTCATCGTAAACGCTAACAAACGGTTTCGAATTTC

      4570       4580       4590       4600       4610       4620
CAGAATATATATTTAATCCATTTCGATCATTCGTAAAGAGTAACATGCAACAAGCTGTAA
GTCTTATATATAAATTAGGTAAAGCTAGTAAGCATTTCTCATTGTACGTTGTTCGACATT

      4630       4640       4650       4660       4670       4680
AAAACATCGATTGTAGTATATATGCACATGGTTGGTTTGGAACCAGATCGAGAGATAATC
TTTTGTAGCTAACATCATATATACGTGTACCAACCAAACCTTGGTCTAGGTCTCTATTAG

      4690       4700       4710       4720       4730       4740
GCGTCGACCAGGTCAGTTGGGGTCAATGATTTTCGCATTAGGAGGCCTAATTTCTGACCA
CGCAGCTGGTCCAGTCAACCCCAGTTACTAAAAGCGTAATCCTCCGGATTAAAGACTGGT
                                                <*  N  R  V  V
                                                <_____________
```

FIG. 4G

```
      4750       4760       4770       4780       4790       4800
CGGAGCAGTACTCCCGCAGAGCCTTCATATCACAGGACTTTTTGCAGCACCTCTCCACGA
GCCTCGTCATGAGGGCGTCTCGGAAGTATAGTGTCCTGAAAAACGTCGTGGAGAGGTGCT
  <S  C  Y  E  R  L  A  K  M  D  C  S  K  K  C  C  R  E  V  I
<_____________DINS4 CODING REGION (EXON-2)_____________

      4810       4820       4830       4840       4850       4860
TTCCTTGCCGTTGGCGAGTTCGCCTCCGGACTTCAGCCAGGGAATTGAGTACACCCCCAA
AAGGAACGGCAACCGCTCAAGCGGAGGCCTGAAGTCGGTCCCTTAACTCATGTGGGGGTT
  <G  Q  R  Q  R  T  R  R  R  V  E  A  L  S  N  L  V  G  G  L
<_____________DINS4 CODING REGION (EXON-2)_____________

      4870       4880       4890       4900       4910       4920
GATAGCTCCCAGGAAAGAGGGCACTTCGCAGCGGTTCCGATATCGAGTTATCCTCCTCCT
CTATCGAGGGTCCTTTCTCCCGTGAAGCGTCGCCAAGGCTATAGCTCAATAGGAGGAGGA
  <Y  S  G  P  F  L  A  S  R  L  P  E  S  I  S  N  D  E  E  E
<_____________DINS4 CODING REGION (EXON-2)_____________

      4930       4940       4950       4960       4970       4980
CGAACTCCTGGACAAACTGCAGGGGATTGAGGGCGTCCAGATCGCTGTCGGCACCGGCTA
GCTTGAGGACCTGTTTGACGTCCCCTAACTCCCGCAGGTCTAGCGACAGCCGTGGCCGAT
  <F  E  Q  V  F  Q  L  P  N  L  A  D  L  D  S  D  A  G  P
<_____________DINS4 CODING REGION (EXON-2)_____________

      4990       5000       5010       5020       5030       5040
ATAAAAATCGTGGATACAATGTAGATCTAGCAAAGCCAGCTTGAGGATCTGCATCCTTGT
TATTTTTAGCACCTATGTTACATCTAGATCGTTTCGGTCGAACTCCTAGACGTAGGAACA

      5050       5060       5070       5080       5090       5100
AAGAACTTACGCATGGCGCGCTTGTGTGGAATCACGGGATTATACTCCTCGCACACCATA
TTCTTGAATGCGTACCGCGCGAACACACCTTAGTGCCCTAATATGAGGAGCGTGTGGTAT
            <M  A  R  K  H  P  I  V  P  N  Y  E  E  C  V  M
          <_____________DINS4 CODING REGION (EXON 1)_____________

      5110       5120       5130       5140       5150       5160
CTCAGCACCTCGTTGAGCTTTTCACTGCAGAGCGTTCCTTGGGCCAACTTCACTGTGGAG
GAGTCGTGGAGCAACTCGAAAAGTGACGTCTCGCAAGGAACCCGGTTGAAGTGACACCTC
<S  L  V  E  N  L  K  E  S  C  L  T  G  Q  A  L  K  V  T  S
<_____________DINS4 CODING REGION (EXON 1)_____________

      5170       5180       5190       5200       5210       5220
CTGGCCAGCAAAATCACGGCCACCATCGAGATGAAGGACAAAGGCTTGCTCATGGTTATG
GACCGGTCGTTTTAGTGCCGGTGGTAGCTCTACTTCCTGTTTCCGAACGAGTACCAATAC
<S  A  L  L  I  V  A  V  M  S  I  F  S  L  P  K  S  M
<_____________DINS4 CODING REGION (EXON 1)_____________

      5230       5240       5250       5260       5270       5280
GGTTTACTGCTTAGGTTGCTTTACGATCAAATGGATTAAGTTGGGTCGAGCCGGGTCGAA
CCAAATGACGAATCCAACGAAATGCTAGTTTACCTAATTCAACCCAGCTCGGCCCAGCTT
```

FIG. 4H

```
      5290       5300       5310       5320       5330       5340
AGCTAACTGATGATGTTTGGCCCAAAGTAACTGGCTTATATACTGCCTCGTAAGAAACTT
TCGATTGACTACTACAAACCGGGTTTCATTGACCGAATATATGACGGAGCATTCTTTGAA
      <__________DINS4 PUT. PROMOTER___________

      5350       5360       5370       5380       5390       5400
AAACTGGGTCTGGGTCGGGGTCGGTCTCTCGGGGTCGGGGTCTGGATCCACACACATGTT
TTTGACCCAGACCCAGCCCCAGCCAGAGAGCCCCAGCCCCAGACCTAGGTGTGTGTACAA

      5410       5420       5430       5440       5450       5460
ATCCTCAAAAGTCAGGTTGTCAAATTGTGTTAGGATGCGATGAGTGCATTCCGGAGTTGG
TAGGAGTTTTCAGTCCAACAGTTTAACACAATCCTACGCTACTCACGTAAGGCCTCAACC

      5470       5480       5490       5500       5510       5520
CTCTTCTCTCTAACGCCTGGCTAAACTCATTCAATGTCAAAGCTGACTTATGCAAATGGC
GAGAAGAGAGATTGCGGACCGATTTGAGTAAGTTACAGTTTCGACTGAATACGTTTACCG

      5530       5540       5550       5560       5570       5580
TATTGGAAAATTGTGGGTGGTTTTTGGGTGGCTGTGTTTGGGAGAAGAAGGGCTTTGTGG
ATAACCTTTTAACACCCACCAAAAACCCACCGACACAAACCCTCTTCTTCCCGAAACACC

      5590       5600       5610       5620       5630       5640
GCGTTTTGCTGTCAGCCAATTAAACAATTTATGTATAAACAGCCAGGCCGTACTAAGCCC
CGCAAAACGACAGTCGGTTAATTTGTTAAATACATATTTGTCGGTCCGGCATGATTCGGG

      5650       5660       5670       5680       5690       5700
TGCATTTATGAATACCAAATAAGTCCTTGGTCTTAAAGTTACCTCGCCTTTACAGCCCGT
ACGTAAATACTTATGGTTTATTCAGGAACCAGAATTTCAATGGAGCGGAAATGTCGGGCA

      5710       5720       5730       5740       5750       5760
TTGCCTCTACCATTTCTACCCTATACTTACCAATCCGCGCCTGGGCGCCCGGCAGGCCGG
AACGGAGATGGTAAAGATGGGATATGAATGGTTAGGCGCGGACCCGCGGGCCGTCCGGCC

      5770       5780       5790       5800       5810       5820
AGTAGGCCAACAAGAACCCGAGCCAGCTGATTGGAGCCAGCAGCATCCTGGCAACGAATT
TCATCCGGTTGTTCTTGGGCTCGGTCGACTAACCTCGGTCGTCGTAGGACCGTTGCTTAA

      5830       5840       5850       5860       5870       5880
ACGCCTCCTTGGTACTTTTCCTTTGACTGTCTTGTCTTTGCCGCTCACACAAATTCTTCT
TGCGGAGGAACCATGAAAAGGAAACTGACAGAACAGAAACGGCGAGTGTGTTTAAGAAGA

          <dIns3_put._poly(A)_signal
                    |
      5890       5900 |     5910       5920       5930       5940
TTTTGCACTGTCTACTTTTATTCATTAGTCAAAGTTGGTGCTGCATAAATAAGTGATTAC
AAAACGTGACAGATGAAAATAAGTAATCAGTTTCAACCACGACGTATTTATTCACTAATG

      5950       5960       5970       5980       5990       6000
GAATTGGATTACGAATGCTGTTAGGAGAACGGGTGTACATATAGTATGTATGTGGGAATG
CTTAACCTAATGCTTACGACAATCCTCTTGCCCACATGTATATCATACATACACCCTTAC
```

FIG. 4I

```
      6010       6020       6030       6040       6050       6060
CCATGTTCAAGTGTTCGTATGTATGTATGTATGTGTATGCATGCTGGGTAATGAATGTGT
GGTACAAGTTCACAAGCATACATACATACATACACATACGTACGACCCATTACTTACACA

      6070       6080       6090       6100       6110       6120
GTGTGTTGGCCAAGTGTCCTATTTCGGTAGACAGTAGATGGCTAACTCCAAGTAGCTGCA
CACACAACCGGTTCACAGGATAAAGCCATCTGTCATCTACCGATTGAGGTTCATCGACGT
                   <*  K  P  L  C  Y  I  A  L  E  L  Y  S  C
                  <___________DINS3 CODING REGION___________

      6130       6140       6150       6160       6170       6180
GGTCTTGACGCAGCACTCGTCGTAGACGCCACCGGTCAGGTGTCTCCTGTGGCGACGCAT
CCAGAACTGCGTCGTGAGCAGCATCTGCGGTGGCCAGTCCACAGAGGACACCGCTGCGTA
<T  K  V  C  C  E  D  Y  V  G  G  T  L  H  R  R  H  R  R  M
<_____________________DINS3 CODING REGION_____________________

      6190       6200       6210       6220       6230       6240
CTTGATCAGGACCTCGGATCCGTACAGATTGGTTAGCAGTGGACTAAAAGAGTATCCTGC
GAACTAGTCCTGGAGCCTAGGCATGTCTAACCAATCGTCACCTGATTTTCTCATAGGACG
<K  I  L  V  E  S  G  Y  L  N  T  L  L  P  S  F  S  Y  G  A
<_____________________DINS3 CODING REGION_____________________

      6250       6260       6270       6280       6290       6300
CCCGTCCAGTGTCTGCCACATGCTGCTATCATCCTGCACCTCCTGCTCCGTGTCCTCGTC
GGGCAGGTCACAGACGGTGTACGACGATAGTAGGACGTGGAGGACGAGGCACAGGAGCAG
<G  D  L  T  Q  W  M  S  S  D  D  Q  V  E  Q  E  T  D  E  D
<_____________________DINS3 CODING REGION_____________________

      6310       6320       6330       6340       6350       6360
GTCGTCGCTGTTGCCCAGCAAGCTTTCACGTTTCCTTGGCAGCGTATTAAAGCCATGGGG
CAGCAGCGACAACGGGTCGTTCGAAAGTGCAAAGGAACCGTCGCATAATTTCGGTACCCC
<D  D  S  N  G  L  L  S  E  R  K  R  P  L  T  N  F  G  H  P
<_____________________DINS3 CODING REGION_____________________

      6370       6380       6390       6400       6410       6420
ACACACCACATCCATGGCATCGGACAGTGCGGGGCCGCAGAGTTTGTGGTTTCCGGGGGG
TGTGTGGTGTAGGTACCGTAGCCTGTCACGCCCCGGCGTCTCAAACACCAAAGGCCCCCC
<C  V  V  D  M  A  D  S  L  A  P  G  C  L  K  H  N  G  P  P
<_____________________DINS3 CODING REGION_____________________

      6430       6440       6450       6460       6470       6480
CAGCAACTGGTGACCACTGCCAGTCGGCGTGACCATTGCCATTGCAGCGGTGAGCATGGC
GTCGTTGACCACTGGTGACGGTCAGCCGCACTGGTAACGGTAACGTCGCCACTCGTACCG
<L  L  Q  H  G  S  G  T  P  T  V  M  A  M  A  A  T  L  M  A
<_____________________DINS3 CODING REGION_____________________

      6490       6500       6510       6520       6530       6540
TGCGATGAGCAGCGACTGGAGCCGAAGGCCATGTACTGCTGCACCGTTGTGCTGGCTAAA
ACGCTACTCGTCGCTGACCTCGGCTTCCGGTACATGACGACGTGGCAACACGACCGATTT
<A  I  L  L  S  Q  L  R  L  G  H  V  A  A  G  N  H  Q  S  F
<_____________________DINS3 CODING REGION_____________________
```

FIG. 4J

```
      6550       6560       6570       6580       6590       6600
CATCTTGGATATGCAGTGAATGCTCTGGGCTGCAACTGGTATTTATACCACTAAATCGGA
GTAGAACCTATACGTCACTTACGAGACCCGACGTTGACCATAAATATGGTGATTTAGCCT
<M
<____

                                                  <__________

      6610       6620       6630       6640       6650       6660
AGCGCTAGCTAATGCAGTTCAATGGCCTCTTCTGCAGTCTAGCATTGCAGTGGCATAGCA
TCGCGATCGATTACGTCAAGTTACCGGAGAAGACGTCAGATCGTAACGTCACCGTATCGT
<__________________DINS3 PUT. PROMOTER____________________

      6670       6680       6690       6700       6710       6720
AGCCCCACGGGCGTACAAACTGCAAATCCTTTGATCACCCATGTTTCAGGTACCGTTTTT
TCGGGGTGCCCGCATGTTTGACGTTTAGGAAACTAGTGGGTACAAAGTCCATGGCAAAAA

      6730       6740       6750       6760       6770       6780
CCCCTAAAAATGCAAACTCTATTTCTAGCTCTACTCCCCAATTTGGATGGAAAAGCGATG
GGGGATTTTTACGTTTGAGATAAAGATCGAGATGAGGGGTTAAACCTACCTTTTCGCTAC

      6790       6800       6810       6820       6830       6840
CACTGTTGTTTTGGTAGTTGGGGTATTGTATTGTATTTCTTAGCAAATATCAGTTGTATC
GTGACAACAAAACCATCAACCCCATAACATAACATAAAGAATCGTTTATAGTCAACATAG

      6850       6860       6870       6880       6890       6900
ATTACCTATATCTATCTATACCAATAGTTTGGAATGTATTTGTAAGACATTTTTAAGATA
TAATGGATATAGATAGATATGGTTATCAAACCTTACATAAACATTCTGTAAAAATTCTAT

      6910       6920       6930       6940       6950       6960
TTCAGAAGAGTTAGCCTTATGGGACTTGCTCTAAAGTGTGAATTGATGCACACAGCTTTA
AAGTCTTCTCAATCGGAATACCCTGAACGAGATTTCACACTTAACTACGTGTGTCGAAAT

      6970       6980       6990       7000       7010       7020
TCGAGCATAGTTTTCAGTGTAATCACCGCCAAAAAATCCGCCCACTTCAAAGCATAACCC
AGCTCGTATCAAAAGTCACATTAGTGGCGGTTTTTTAGGCGGGTGAAGTTTCGTATTGGG

      7030       7040       7050       7060       7070       7080
GTTCGCCCAACCTGTTACATTGCCGCTAAGAGGCTCTGACTGCTGTCGATTGCGATTACG
CAAGCGGGTTGGACAATGTAACGGCGATTCTCCGAGACTGACGACAGCTAACGCTAATGC

      7090       7100       7110       7120       7130       7140
ATTACGACCAGATATCTGTGGGGCATGGGGATAAGGGGTATGTGGGCCGATGGCTGACAG
TAATGCTGGTCTATAGACACCCCGTACCCCTATTCCCCATACACCCGGCTACCGACTGTC

      7150       7160       7170       7180       7190       7200
TGTGGCAGCCTCATTAGCATGTCGTGGCCAGGAGGAAAGTATGCTTCGATGAAGCTCCTC
ACACCGTCGGAGTAATCGTACAGCACCGGTCCTCCTTTCATACGAAGCTACTTCGAGGAG

      7210       7220       7230       7240       7250       7260
CGGCGGCAGTGTGCGAAATCGCTTCGATCACCATCATCGCCATCGCCATGGCCACTCGAT
GCCGCCGTCACACGCTTTAGCGAAGCTAGTGGTAGTAGCGGTAGCGGTACCGGTGAGCTA
```

FIG. 4K

```
      7270       7280       7290       7300       7310       7320
TGTCGAGTTGCACGCACGGCGATGCCAACAGTTGGTTGCCAGCGCTGCACTCGAAACACT
ACAGCTCAACGTGCGTGCCGCTACGGTTGTCAACCAACGGTCGCGACGTGAGCTTTGTGA

      7330       7340       7350       7360       7370       7380
CGCTTCTTCCCACCGACCGCAAAGTGCCGGAAAAGCTAGAAAAAAAGCAAAAAAAAAAGT
GCGAAGAAGGGTGGCTGGCGTTTCACGGCCTTTTCGATCTTTTTTTCGTTTTTTTTTTCA

      7390       7400       7410       7420       7430       7440
GGAAGAAAATTCGCGATAGAAAACGGAAAAATCGAAACGAACAAAAAAAGTCGGAATAAA
CCTTCTTTTAAGCGCTATCTTTTGCCTTTTTAGCTTTGCTTGTTTTTTTCAGCCTTATTT

      7450       7460       7470       7480       7490       7500
TCAAGGAAACATGGTGCTCGACATTAAGATGTGCCGATTTGATAATGTGCCCTGGGGCTT
AGTTCCTTTGTACCACGAGCTGTAATTCTACACGGCTAAACTATTACACGGGACCCCGAA

      7510       7520       7530       7540       7550       7560
TCGCCTGGTGGGCGGGGCGGACTACGATTATCCGCTGACGGTGGTTAAGGTTAGGCCCGA
AGCGGACCACCCGCCCCGCCTGATGCTAATAGGCGACTGCCACCAATTCCAATCCGGGCT

      7570       7580       7590       7600       7610       7620
TTCGAAAAAAGAACGAAATCTATATGCTGCAACCCCCACCCCCCCACGCATCACCTCAGC
AAGCTTTTTTCTTGCTTTAGATATACGACGTTGGGGGGTGGGGGGGGTGCGTAGTGGAGTCG

      7630       7640       7650       7660       7670       7680
CCATTCACCTGGCGGATGTTCATAGACCAGTGGAAAATATTGCTCACTATGCAGCTGATG
GGTAAGTGGACCGCCTACAAGTATCTGGTCACCTTTTATAACGAGTGATACGTCGACTAC

      7690       7700       7710       7720       7730       7740
AATCACATTGGATTAATTCGATACGATACGTTCGAATCAGTTTTATTTGTTCGATTGCAA
TTAGTGTAACCTAATTAAGCTATGCTATGCAAGCTTAGTCAAAATAAACAAGCTAACGTT

      7750       7760       7770       7780       7790       7800
TATTACGTAACGCCGCGATGCGTGTGTGTCCATTCGGATTTGCTGCATTGGCAAATTAGT
ATAATGCATTGCGGCGCTACGCACACACAGGTAAGCCTAAACGACGTAACCGTTTAATCA

      7810       7820       7830       7840       7850       7860
TAATTAAAGTAATTCCTCTCGCTTTTGTTTATCTAATCGACAGGGCCATACATTTCCCGC
ATTAATTTCATTAAGGAGAGCGAAAACAAATAGATTAGCTGTCCCGGTATGTAAAGGGCG

      7870       7880       7890       7900       7910       7920
TAATGAGCCGCATAATGGCAGCGGCAATAAACTTATTCAAATTTTAATTGTGTTTCGCTG
ATTACTCGGCGTATTACCGTCGCCGTTATTTGAATAAGTTTAAAATTAACACAAAGCGAC

      7930       7940       7950       7960       7970       7980
GCAGTTGGTCCTTTGTTTGTGCATAAATTGCATTTGGCAATTCGCATTTTGTAACATTGT
CGTCAACCAGGAAACAAACACGTATTTAACGTAAACCGTTAAGCGTAAAACATTGTAACA

      7990       8000       8010       8020       8030       8040
GTTGACAATTCGCAACCAGCAACAATAACAAAAATACAATACATACAATACTATAGCATC
CAACTGTTAAGCGTTGGTCGTTGTTATTGTTTTTATGTTATGTATGTTATGATATCGTAG
```

FIG. 4L

```
      8050       8060       8070       8080       8090       8100
GTCGTAAATCCGAAACAAATGCGATTTTTAATTGGCAAACTGCTAAGCGCATAAAACAAA
CAGCATTTAGGCTTTGTTTACGCTAAAAATTAACCGTTTGACGATTCGCGTATTTTGTTT

      8110       8120       8130       8140       8150       8160
TGACCGAAATGCGAGGGGCGCTAAAAAATCCCATCCCTTCGATACGAATAAATCAATTTA
ACTGGCTTTACGCTCCCCGCGATTTTTTAGGGTAGGGAAGCTATGCTTATTTAGTTAAAT

      8170       8180       8190       8200       8210       8220
AGCCGCAGAGTCAAGGAAGGGAGGTCATAAATTGTTTTTGACTTTTTGGTTATTTTTTTT
TCGGCGTCTCAGTTCCTTCCCTCCAGTATTTAACAAAAACTGAAAAACCAATAAAAAAAA

      8230       8240       8250       8260       8270       8280
TTACCGTTTTACATAAACAAATTATGCTATGGGTTATTTTAAATTCCGATCAATTTATAA
AATGGCAAAATGTATTTGTTTAATACGATACCCAATAAAATTTAAGGCTAGTTAAATATT

      8290       8300       8310       8320       8330       8340
AATGTTTGTGCTTTGGGATATGCATACCATGAAAAAATGGAAGTTTATTGTAAATGAATT
TTACAAACACGAAACCCTATACGTATGGTACTTTTTTACCTTCAAATAACATTTACTTAA

      8350       8360       8370       8380       8390       8400
ATTAACTTCACAAGCTGGCTGATAGAGAAAAAACTGAAAAATGTCCGGAATGTTCTTCAT
TAATTGAAGTGTTCGACCGACTATCTCTTTTTTGACTTTTTACAGGCCTTACAAGAAGTA

      8410       8420       8430       8440       8450       8460
TCCAATGAACTCCCTAAATTAACTTAGCTAATTTATTCCTTATACTAATACTCCGCTTTT
AGGTTACTTGAGGGATTTAATTGAATCGATTAAATAAGGAATATGATTATGAGGCGAAAA

      8470       8480       8490       8500       8510       8520
AAGAATTCCTTACTACATGTTAGAGACTCAAAAAGCACATCCTTCGACTCGAGTCCATAT
TTCTTAAGGAATGATGTACAATCTCTGAGTTTTTCGTGTAGGAAGCTGAGCTCAGGTATA

      8530       8540       8550       8560       8570       8580
TACTTTATGGAATGTGCCAACACACCTTCACATATTGGCTCTGCAAACACTAAACAATCC
ATGAAATACCTTACACGGTTGTGTGGAAGTGTATAACCGAGACGTTTGTGATTTGTTAGG

      8590       8600       8610       8620       8630       8640
TTGGTAATCTTTTGAAAAACCTCTGTTTACACTACCACTCTTCGTCATGCTGCTCGCCAC
AACCATTAGAAAACTTTTTGGAGACAAATGTGATGGTGAGAAGCAGTACGACGAGCGGTG

      8650       8660       8670       8680       8690       8700
ATACAGTCTGGTACATAGATGTATGGCCCAGCTAAGCCCAAAGCCTTTGTTCTATAAATA
TATGTCAGACCATGTATCTACATACCGGGTCGATTCGGGTTTCGGAAACAAGATATTTAT

      8710       8720       8730       8740       8750       8760
TTCGCAACCTCCGACGATGTCGAGTGCTTTTTGCTCTGCGAATTCACCGCTGGAAATTGA
AAGCGTTGGAGGCTGCTACAGCTCACGAAAAACGAGACGCTTAAGTGGCGAGCTTTAACT

      8770       8780       8790       8800       8810       8820
CTCTACCATAAGTGAAATGCAAGAGACCCCTGGGACTGAAAGGAAAGACCCTCAACTTGG
GAGATGGTATTCACTTTACGTTCTCTGGGGACCCTGACTTTCCTTTCTGGGAGTTGAACC
```

FIG. 4M

```
      8830       8840       8850       8860       8870       8880
TTGGGTGAAATGGTGGAGTCTCCAACCCTCCACCTGCTCCTTGTGCCAACCACTTTTTTT
AACCCACTTTACCACCTCAGAGGTTGGGAGGTGGACGAGGAACACGGTTGGTGAAAAAAA

      8890       8900       8910       8920       8930       8940
TTTTTGCAGTATTTGCATTACTAAGTCCTCTTGGCAGTCGGTGTCGTGACTTTCTGGTTA
AAAAACGTCATAAACGTAATGATTCAGGAGAACCGTCAGCCACAGCACTGAAAGACCAAT

      8950       8960       8970       8980       8990       9000
TGAACCCTGCTTTCTCATAACGGAAACGAAAACAATCGCGTTTATTTGCCCACGAAAGTG
ACTTGGGACGAAAGAGTATTGCCTTTGCTTTTGTTAGCGCAAATAAACGGGTGCTTTCAC

      9010       9020       9030       9040       9050       9060
TTACAAAACTGCCTGAATTATGCAATAGAATTCTTTGAACAGAGTGCTAAGATATTTCGC
AATGTTTTGACGGACTTAATACGTTATCTTAAGAAACTTGTCTCACGATTCTATAAAGCG

      9070       9080       9090       9100       9110       9120
ATTTTGAAGGCGAAACATAATTCATCCATAACTATTAGTTTGATGAATTCTCACTTCGTA
TAAAACTTCCGCTTTGTATTAAGTAGGTATTGATAATCAAACTACTTAAGAGTGAAGCAT

      9130       9140       9150       9160       9170       9180
TGCTGGCAATTTTGAAGGCCGAAGTGGCAAAACCATTTTAAATGAATTCCTACAATTTAT
ACGACCGTTAAAACTTCCGGCTTCACCGTTTTGGTAAAATTTACTTAAGGATGTTAAATA

      9190       9200       9210       9220       9230       9240
ATGCTCAATTTGCGCCAATTTGTCTGTACTTTATTACCCACAAAAGCCATAAAGCTTATA
TACGAGTTAAACGCGGTTAAACAGACATGAAATAATGGGTGTTTTCGGTATTTCGAATAT

      9250       9260       9270       9280       9290       9300
TTGTATTTAGTTGTTATTGTTCTCGCAACATTTTCACTTTGATTTATAGTTGCGAAAATA
AACATAAATCAACAATAACAAGAGCGTTGTAAAAGTGAAACTAAATATCAACGCTTTTAT

      9310       9320       9330       9340       9350       9360
AATGTTGGCAAATGCAAATTTTTGAAACGATATTCTCTTACACGGTCATTTGGTACCATT
TTACAACCGTTTACGTTTAAAAACTTTGCTATAAGAGAATGTGCCAGTAAACCATGGTAA

      9370       9380       9390       9400       9410       9420
TCCGGAATATACATTTACCAATTTTCCAAAAAAAAGAGCCATAAATTGTATTATCCAATTA
AGGCCTTATATGTAAATGGTTAAAAGGTTTTTTTCTCGGTATTTAACATAATAGGTTAAT

      9430       9440       9450       9460       9470       9480
AAACGATTAGTCCAGTGGGTTTTTATTTCTAAAAATTTAATTTTGTAATTAGAGAAAACT
TTTGCTAATCAGGTCACCCAAAAATAAAGATTTTTAAATTAAAACATTAATCTCTTTTGA

      9490       9500       9510       9520       9530       9540
ATTGTGAACTTTGGATGACCTAGAAAGTTTCATTAGTTGTACATTATTTTTTACCCCCGC
TAACACTTGAAACCTACTGGATCTTTCAAAGTAATCAACATGTAATAAAAAATGGGGGCG

      9550       9560       9570       9580       9590       9600
TTTAAAATTGCAGAACTTCTGTAAAAAAAAGCTTTTACAAGCTATTTAAATATTAGTGGT
AAATTTTAACGTCTTGAAGACATTTTTTTTCGAAAATGTTCGATAAATTTATAATCACCA
```

FIG. 4N

```
      9610       9620       9630       9640       9650       9660
AGAGTTGTTTGAACATTTATCTTTTTGGAATCAGCAATATTTGGTTCTCTATCCTTACAA
TCTCAACAAACTTGTAAATAGAAAAACCTTAGTCGTTATAAACCAAGAGATAGGAATGTT

      9670       9680       9690       9700       9710       9720
ATTAATCTTCTCTTTAAGTGAACTCATAACTTTGTATATATGTACGTATGTAGAATCATC
TAATTAGAAGAGAAATTCACTTGAGTATTGAAACATATATACATGCATACATCTTAGTAG

      9730       9740       9750       9760       9770       9780
CTTCAACTCAAACATTTCCATCCGGCGAAACAATGCAATATTTGAGTGGTTAGACATGGG
GAAGTTGAGTTTGTAAAGGTAGGCCGCTTTGTTACGTTATAAACTCACCAATCTGTACCC

      9790       9800       9810       9820       9830       9840
TAATTCCACTTGACTCGCTTAACTGAAGTGTCGTTAATGAGCTGCCACTTCTACTCGAGC
ATTAAGGTGAACTGAGCGAATTGACTTCACAGCAATTACTCGACGGTGAAGATGAGCTCG

      9850       9860       9870       9880       9890       9900
ACCCCTCGTTCTGAGCCCCCAGCCCCCACAGATCCTCTGTAGCCCCCCCATCTCCTTGGG
TGGGGAGCAAGACTCGGGGGTCGGGGGTGTCTAGGAGACATCGGGGGGGTAGAGGAACCC

      9910       9920       9930       9940       9950       9960
CATTGTCTTCGGTTCTGTTGGGTTTGTGCGTGATTGTATGATTGTCTGTTCGGGGGTCTG
GTAACAGAAGCCAAGACAACCCAAACACGCACTAACATACTAACAGACAAGCCCCCAGAC

      9970       9980       9990      10000      10010      10020
GGTCTGCTGGTCTTTGTTTTGTTGACATTTGGCGCGCGTTTTATTTTTATGCACTTAGCA
CCAGACGACCAGAAACAAAACAACTGTAAACCGCGCGCAAAATAAAAATACGTGAATCGT

     10030      10040      10050      10060      10070      10080
CGCGACGTCGCCGTTGTCGGCTAATAGAAATTTCCCCATTATCGCATCGCATCCCATTGT
GCGCTGCAGCGGCAACAGCCGATTATCTTTAAAGGGGTAATAGCGTAGCGTAGGGTAACA

     10090      10100      10110      10120      10130      10140
ATTGTATCTCGGCTATCTCGACTATCTCGGCCTCGCACTCATTCTATCGCCACATTCCCA
TAACATAGAGCCGATAGAGCTGATAGAGCCGGAGCGTGAGTAAGATAGCGGTGTAAGGGT

     10150      10160      10170      10180      10190      10200
TATCCGCATCTGAATCGGGCGATATCTAGGCATTCCCATCTAGATCTAAACATGTCCATA
ATAGGCGTAGACTTAGCCCGCTATAGATCCGTAAGGGTAGATCTAGATTTGTACAGGTAT

     10210      10220      10230      10240      10250      10260
TGCTTCAACGTTAGGTGACCGAGGGCAGCATTGCTGACGAGGCTGGACTGCGGGTCGAGG
ACGAAGTTGCAATCCACTGGCTCCCGTCGTAACGACTGCTCCGACCTGACGCCCAGCTCC

     10270      10280      10290      10300      10310      10320
ATATCATCGTGCGCATCAATGACACGGCTGCCACGCCCCTTACCCACGACGAGGCCCACC
TATAGTAGCACGCGTAGTTACTGTGCCGACGGTGCGGGGAATGGGTGCTGCTCCGGGTGG

     10330      10340      10350      10360      10370      10380
GCCTCATTATGGGCAGTGGAAGCGTCTTCTATTTTGGCGTCTACCGGTGAGCTTTCCCAT
CGGAGTAATACCCGTCACCTTCGCAGAAGATAAAACCGCAGATGGCCACTCGAAAGGGTA
```

FIG. 4O

```
     10390      10400      10410      10420      10430      10440
CCTTCTATTCCATACCATTCCGTTTCGTTCTCTTACTGCTCGTGGTCGTGGTCCTCGTCC
GGAAGATAAGGTATGGTAAGGCAAAGCAAGAGAATGACGAGCACCAGCACCAGGAGCAGG

     10450      10460      10470      10480      10490      10500
TTGTCCGGTTCCTTTACTACTACTCTTTGTATCCCATCCACCGAGGACCATTTATCACAG
AACAGGCCAAGGAAATGATGATGAGAAACATAGGGTAGGTGGCTCCTGGTAAATAGTGTC

     10510      10520      10530      10540      10550      10560
CTGAGCGGGCTAACAACCCAAGAACGTTTCCTCATGCCCCTGCTCAAGGTAATCTACTTG
GACTCGCCCGATTGTTGGGTTCTTGCAAAGGAGTACGGGGACGAGTTCCATTAGATGAAC

     10570      10580      10590      10600      10610      10620
TTCAAGAAATTCACAAACCACAAACGACCTCGAGAGAAATGGAAAAAATATGACAAATTT
AAGTTCTTTAAGTGTTTGGTGTTTGCTGGAGCTCTCTTTACCTTTTTTATACTGTTTAAA

     10630      10640      10650      10660      10670      10680
TCGTGATTTAAGAATAAAGTTCTGGAAAAATAAAGCGCTTTCTTAAAAAGTTGTCTGGGT
AGCACTAAATTCTTATTTCAAGACCTTTTTATTTCGCGAAAGAATTTTTCAACAGACCCA

     10690      10700      10710      10720      10730      10740
AAAATGACATTTGGTTAATATATCATAATAGTTAATTTTATTATAATTATAAACTAAGAA
TTTTACTGTAAACCAATTATATAGTATTATCAATTAAAATAATATTAATATTTGATTCTT

     10750      10760      10770      10780      10790      10800
AAGTTAAATTCAAAAACCCACCTAGCCCCATTAGTTTTGAAAATTACCCTACCATTTTAG
TTCAATTTAAGTTTTTGGGTGGATCGGGGTAATCAAAACTTTTAATGGGATGGTAAAATC

     10810      10820      10830      10840      10850      10860
AAGCAATTTATATTATTTGAATTAAGTTTGTATTTCAACTTTTTCGGGTTATGAATAATT
TTCGTTAAATATAATAAACTTAATTCAAACATAAAGTTGAAAAAGCCCAATACTTATTAA

     10870      10880      10890      10900      10910      10920
ATTCTTAGAGTGTCCCCGAAACCAGGGCTCCATCTCAGGTATTCCACGTTACGGAATTAA
TAAGAATCTCACAGGGGCTTTGGTCCCGAGGTAGAGTCCATAAGGTGCAATGCCTTAATT

     10930      10940      10950      10960      10970      10980
TTCTAATGTTCATAAATTCTGCTCACTTTTTGGTCACTTGGATCCATGTGCAGGGAGAAC
AAGATTACAAGTATTTAAGACGAGTGAAAAACCAGTGAACCTAGGTACACGTCCCTCTTG

     10990      11000      11010      11020      11030      11040
GAGGAGGACGCTTACGAGTGCCTAAAGAAGTTTCCCACGAGCGAGGGTTCGTTGACCAAG
CTCCTCCTGCGAATGCTCACGGATTTCTTCAAAGGGTGCTCGCTCCCAAGCAACTGGTTC

     11050      11060      11070      11080      11090      11100
TCACCAATGCCGACCATTTCACCGTCGCCGACTCCATCGCTGTCCCAGCTGACGGAAACC
AGTGGTTACGGCTGGTAAAGTGGCAGCGGCTGAGGTAGCGACAGGGTCGACTGCCTTTGG

     11110      11120
ACAAATGCCCGTACTCCGGA
TGTTTACGGGCATGAGGCCT
```

FIG. 4P

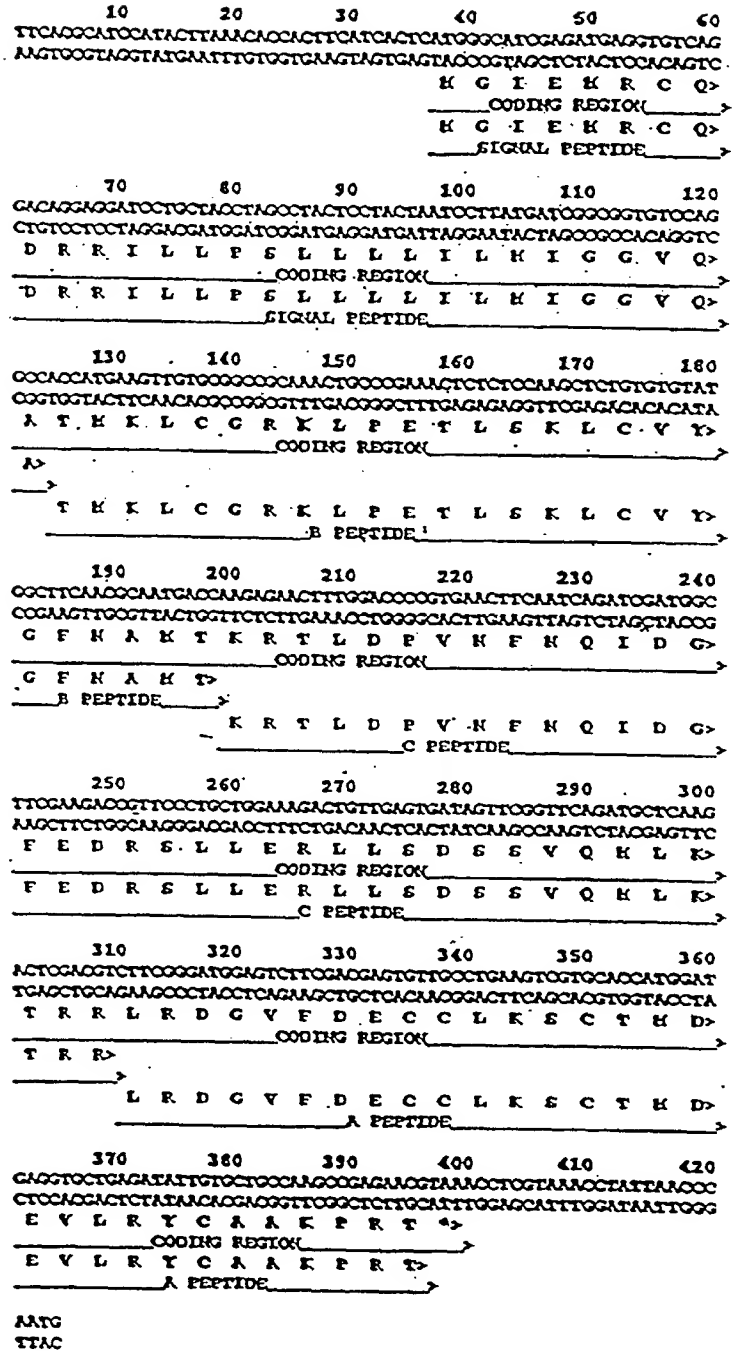

FIG. 5

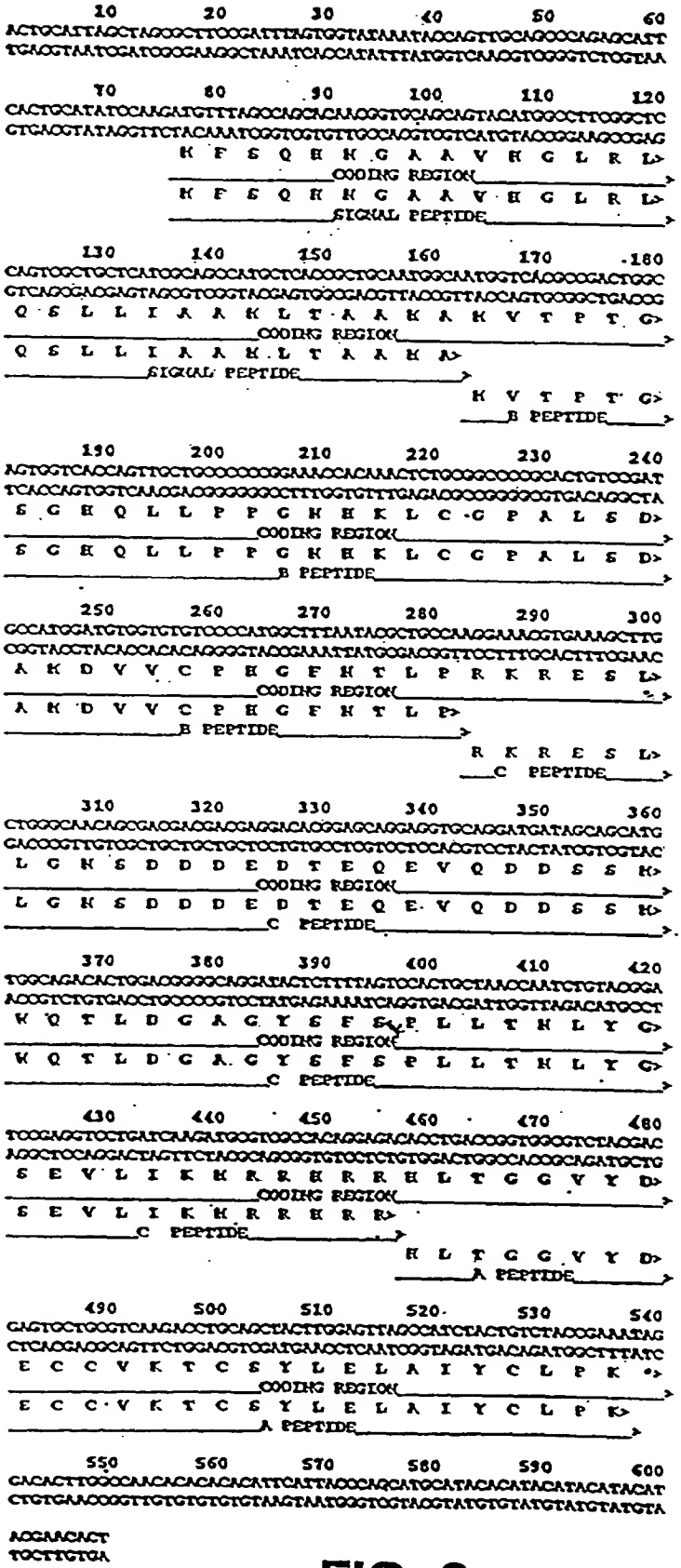

FIG. 6

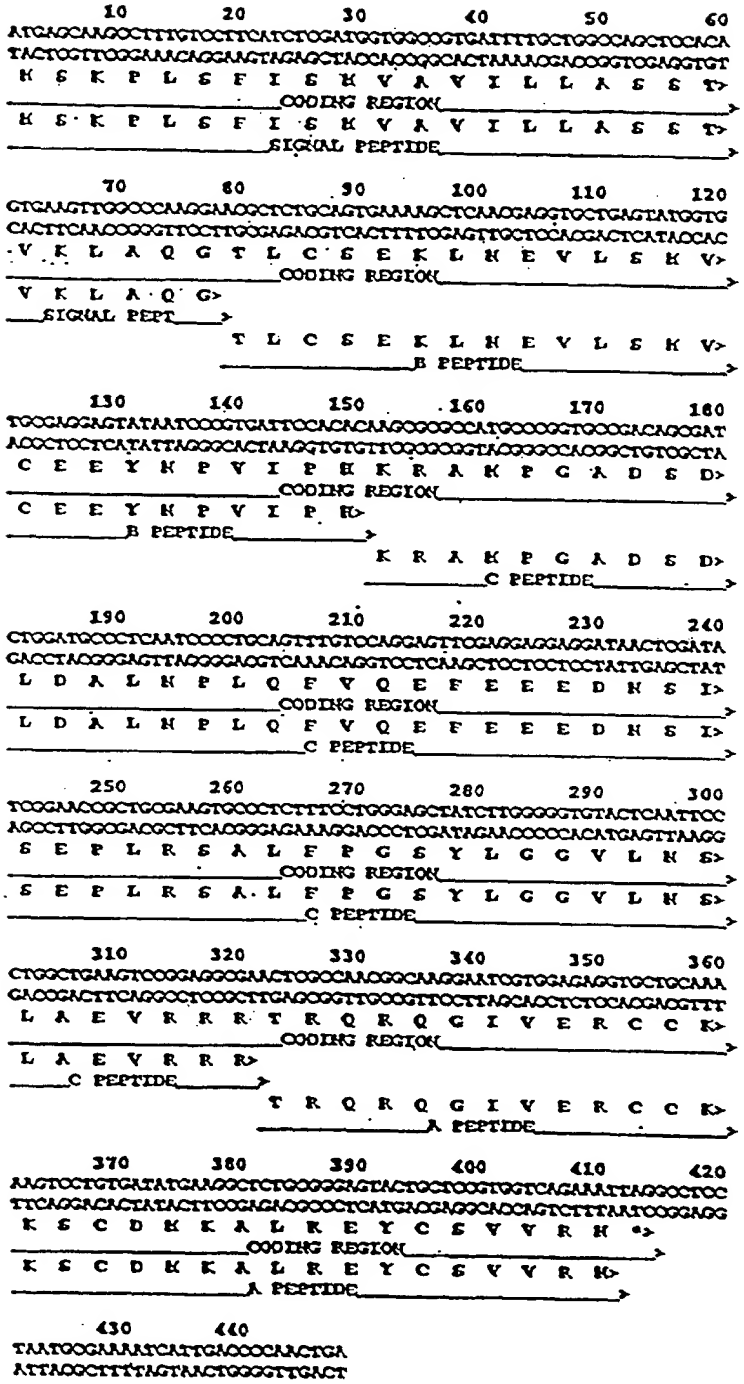

FIG. 7

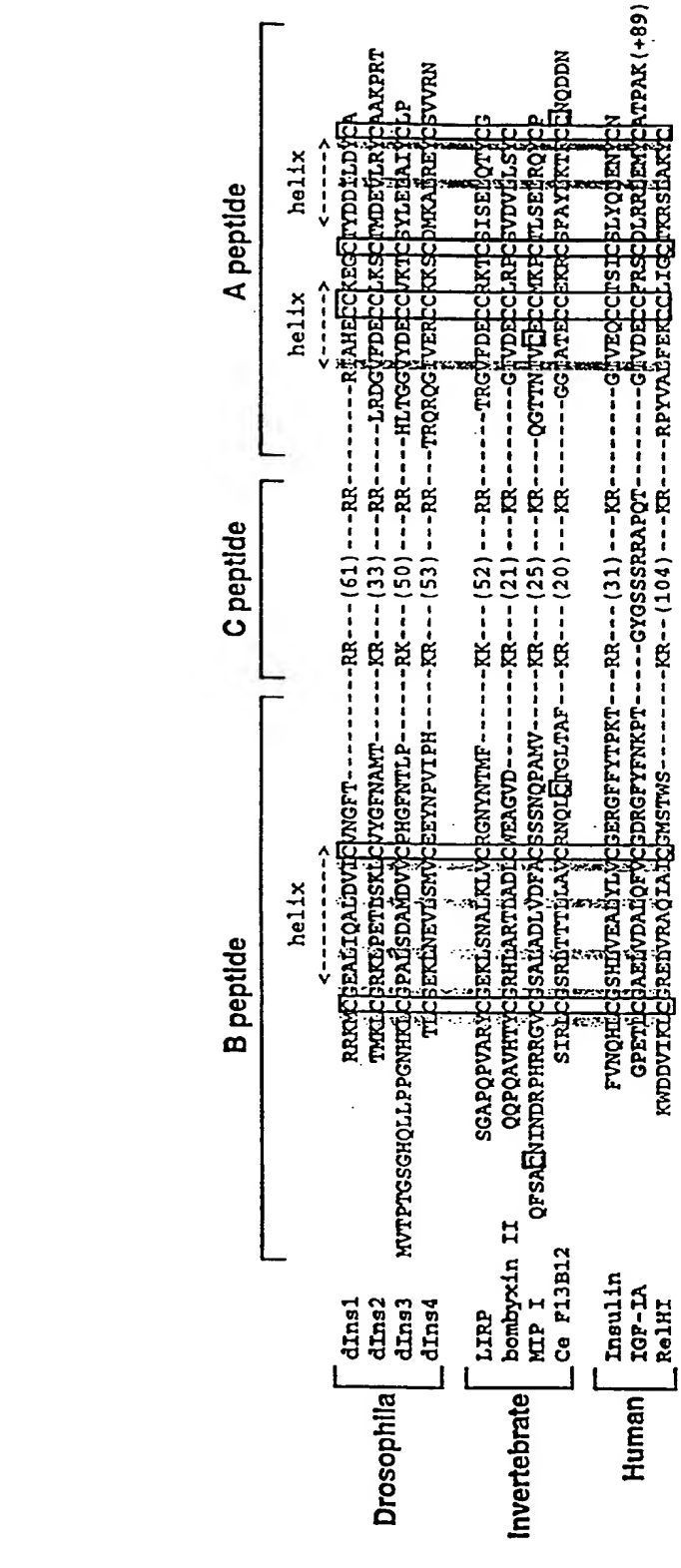

FIG. 8

SEQUENCE LISTING

<110> Exelixis Pharmaceuticals, Inc.

<120> NUCLEIC ACIDS AND PROTEINS OF D. MELANOGASTER
INSULIN-LIKE GENES AND USES THEREOF

<130> 7326-066-228

<140>
<141>

<150> 09/201,227
<151> 1998-11-30

<160> 15

<170> PatentIn Ver. 2.0

<210> 1
<211> 424
<212> DNA
<213> Drosophila melanogaster

<400> 1
ttcacgcatc catacttaaa caccacttca tcactcatgg gcatcgagat gaggtgtcag 60
gacaggagga tcctgctacc tagcctactc ctactaatcc ttatgatcgg cggtgtccag 120
gccaccatga agttgtgcgg ccgcaaactg cccgaaactc tctccaagct ctgtgtgtat 180
ggcttcaacg caatgaccaa gagaactttg gaccccgtga acttcaatca gatcgatggc 240
ttcgaagacc gttccctgct ggaaagactg ttgagtgata gttcggttca gatgctcaag 300
actcgacgtc ttcgggatgg agtcttcgac gagtgttgcc tgaagtcgtg caccatggat 360
gaggtgctga gatattgtgc tgccaagccg agaacgtaaa cctcgtaaac ctattaaccc 420
aatg 424

<210> 2
<211> 120
<212> PRT
<213> Drosophila melanogaster

<400> 2
Met Gly Ile Glu Met Arg Cys Gln Asp Arg Arg Ile Leu Leu Pro Ser
1 5 10 15

Leu Leu Leu Leu Ile Leu Met Ile Gly Gly Val Gln Ala Thr Met Lys
20 25 30

Leu Cys Gly Arg Lys Leu Pro Glu Thr Leu Ser Lys Leu Cys Val Tyr
35 40 45

Gly Phe Asn Ala Met Thr Lys Arg Thr Leu Asp Pro Val Asn Phe Asn
50 55 60

Gln Ile Asp Gly Phe Glu Asp Arg Ser Leu Leu Glu Arg Leu Leu Ser
65 70 75 80

Asp Ser Ser Val Gln Met Leu Lys Thr Arg Arg Leu Arg Asp Gly Val
85 90 95

```
Phe Asp Glu Cys Cys Leu Lys Ser Cys Thr Met Asp Glu Val Leu Arg
            100                 105                 110

Tyr Cys Ala Ala Lys Pro Arg Thr
        115                 120
```

<210> 3
<211> 609
<212> DNA
<213> Drosophila melanogaster

<400> 3
actgcattag ctagcgcttc cgatttagtg gtataaatac cagttgcagc ccagagcatt 60
cactgcatat ccaagatgtt tagccagcac aacggtgcag cagtacatgg ccttcggctc 120
cagtcgctgc tcatcgcagc catgctcacc gctgcaatgg caatggtcac gccgactggc 180
agtggtcacc agttgctgcc ccccggaaac cacaaactct gcggccccgc actgtccgat 240
gccatggatg tggtgtgtcc ccatggcttt aatacgctgc caaggaaacg tgaaagcttg 300
ctgggcaaca gcgacgacga cgaggacacg gagcaggagg tgcaggatga tagcagcatg 360
tggcagacac tggacggggc aggatactct tttagtccac tgctaaccaa tctgtacgga 420
tccgaggtcc tgatcaagat gcgtcgccac aggagacacc tgaccggtgg cgtctacgac 480
gagtgctgcg tcaagacctg cagctacttg gagttagcca tctactgtct accgaaatag 540
gacacttggc caacacacac acattcatta cccagcatgc atacacatac atacatacat 600
acgaacact 609

<210> 4
<211> 154
<212> PRT
<213> Drosophila melanogaster

<400> 4
```
Met Phe Ser Gln His Asn Gly Ala Ala Val His Gly Leu Arg Leu Gln
  1               5                  10                  15

Ser Leu Leu Ile Ala Ala Met Leu Thr Ala Ala Met Ala Met Val Thr
             20                  25                  30

Pro Thr Gly Ser Gly His Gln Leu Leu Pro Pro Gly Asn His Lys Leu
         35                  40                  45

Cys Gly Pro Ala Leu Ser Asp Ala Met Asp Val Val Cys Pro His Gly
     50                  55                  60

Phe Asn Thr Leu Pro Arg Lys Arg Glu Ser Leu Leu Gly Asn Ser Asp
 65                  70                  75                  80

Asp Asp Glu Asp Thr Glu Gln Glu Val Gln Asp Asp Ser Ser Met Trp
                 85                  90                  95

Gln Thr Leu Asp Gly Ala Gly Tyr Ser Phe Ser Pro Leu Leu Thr Asn
            100                 105                 110

Leu Tyr Gly Ser Glu Val Leu Ile Lys Met Arg Arg His Arg Arg His
        115                 120                 125

Leu Thr Gly Gly Val Tyr Asp Glu Cys Cys Val Lys Thr Cys Ser Tyr
    130                 135                 140
```

```
Leu Glu Leu Ala Ile Tyr Cys Leu Pro Lys
145                 150
```

<210> 5
<211> 448
<212> DNA
<213> Drosophila melanogaster

<400> 5
atgagcaagc ctttgtcctt catctcgatg gtggccgtga ttttgctggc cagctccaca 60
gtgaagttgg cccaaggaac gctctgcagt gaaaagctca acgaggtgct gagtatggtg 120
tgcgaggagt ataatcccgt gattccacac aagcgcgcca tgcccggtgc cgacagcgat 180
ctggatgccc tcaatccccct gcagtttgtc caggagttcg aggaggagga taactcgata 240
tcggaaccgc tgcgaagtgc cctctttcct gggagctatc ttgggggtgt actcaattcc 300
ctggctgaag tccggaggcg aactcgccaa cggcaaggaa tcgtggagag gtgctgcaaa 360
aagtcctgtg atatgaaggc tctgcgggag tactgctccg tggtcagaaa ttaggcctcc 420
taatgcgaaa atcattgacc ccaactga 448

<210> 6
<211> 137
<212> PRT
<213> Drosophila melanogaster

<400> 6

```
Met Ser Lys Pro Leu Ser Phe Ile Ser Met Val Ala Val Ile Leu Leu
  1               5                  10                  15

Ala Ser Ser Thr Val Lys Leu Ala Gln Gly Thr Leu Cys Ser Glu Lys
             20                  25                  30

Leu Asn Glu Val Leu Ser Met Val Cys Glu Glu Tyr Asn Pro Val Ile
         35                  40                  45

Pro His Lys Arg Ala Met Pro Gly Ala Asp Ser Asp Leu Asp Ala Leu
     50                  55                  60

Asn Pro Leu Gln Phe Val Gln Glu Phe Glu Glu Glu Asp Asn Ser Ile
 65                  70                  75                  80

Ser Glu Pro Leu Arg Ser Ala Leu Phe Pro Gly Ser Tyr Leu Gly Gly
                 85                  90                  95

Val Leu Asn Ser Leu Ala Glu Val Arg Arg Arg Thr Arg Gln Arg Gln
            100                 105                 110

Gly Ile Val Glu Arg Cys Cys Lys Lys Ser Cys Asp Met Lys Ala Leu
        115                 120                 125

Arg Glu Tyr Cys Ser Val Val Arg Asn
    130                 135
```

<210> 7
<211> 11120
<212> DNA
<213> Drosophila melanogaster

<400> 7
gcttctgctc ggagagcggc tgacccgaat gggatagggc atctcctgtc cacaggaacg 60
aactccccat tatcgccctg cttggccttc atcgtcttca ccgccgattc aatcagtctc 120
caaggactat cgcagttgta gtgtccaaag ggaccaggtg gctccgatgc cgcactattg 180
gcattggagc cggaaaccga tcgtgctaac tcccagcagc ttctgtatat gtcgccctgc 240
gtggagtaga gtgtgtcaag gtggaggtca ctaatgtgcc agaagtagcc tgtcaacaaa 300
caagtagaat caagtaaatg tgttagttaa atacccatag atatatgtaa aagttgttgt 360
tttatttgct aagaaaagtt taatctatat cccagtttta cacaccagat ttttatgtcc 420
tgagcaattt cgtatgtatt tccccttcgt aaagtaagga tcgagattag actttgactt 480
tggttaagtc gggcaattcc tggccgggaa aggccatttc ctttcgcggg gcattttccc 540
gccggctggt cgagcgacaa aaataagaaa aacctggtag ttcaaatgga aatctcctgc 600
agctgactgt ttggttggtt gactgacctg gcccgaattt aactttctac ctggtcgcaa 660
tacgtgaagt caaaaagtca attagcgagt caacattttg agcgccggcc aactccaagg 720
atcagtatca tttggcatgc ccagcgatcg gtttgccaag agcacgagaa gttcgagata 780
ggacccagag ataccagaga taaaggaggc atacctttta tgcccggtga gagcacggac 840
ggcggagtga aagatcgagc aggatgagcc tgattagact gggactggcg ctgctgctcc 900
tgctggccac cgtgtcgcag ttactgcagc cggtccaggg acgccgaaag atgtgcggcg 960
aggctctgat ccaggcactg gatgtgattt gtgttaatgg atttacacgc cgtgtcaggc 1020
ggagcagtgg taagtttggg tactatgcat attcgattgg cttccataca tctaacttct 1080
tttcgacaag cgtctaagga tgctagagtg cgagacctta tccgtaagct acagcagccg 1140
gatgaggaca ttgaacagga aacggaaacg ggaaggttaa agcagaagca tacggatgcg 1200
gatacggaga agggtgtgcc accggccgtc ggaagtggac gaaagttgcg acgccatcgg 1260
cgacgcatcg cccacgagtg ttgcaaggag ggctgcacct acgacgatat actggactac 1320
tgcgcctgat gaccaggatg gcaaaacaaa acaaataaaa accagaaacc agatcccaaa 1380
aaccaagtac cagatgaaca cgacatggct gagattttgt gtggcggcac ggggaaaaca 1440
cccgacgacc ggcaggctat ttgcaattca ttttcctact acacttaacc cctaactata 1500
aacgtaatcg tatttccaaa tatttcattg taaaatttct agtggaggca aataaagtta 1560
ctctccaagc agcagcagaa acaaaagaag agtccattgc ttttttctac attctacgcc 1620
ctgcagcatt ccagctgtga ggcatgggga atccccttgt tattcaaacc acccgaagcc 1680
acccaaacca tcgagccacc cacaagcagc tgccattcag cacctcgagt gcggtgccct 1740
tgttttccga gaacaataat gaaaaatatg aatttttaat tagatgacgt tctgatttta 1800
ataagcaaaa caaaaggtgg agacaaaacg aactcggtaa tacactcaga ttcgaattta 1860
cagcttcctt tttatccata atttttgtta ttatcgaagg agcgatatca aaactagaaa 1920
acaacttcca atcagtagcg ggattttccg aagataacac tctattcaac cgaagggttt 1980
tgaaatgata ataattccgt tcttacaggt aaaaatctat actaatacct gtttttttgc 2040
ggacggaaaa aaggctcagt tggcttatca ttggcaaaag ggacttgggg aaaccataaa 2100
gtatcgaagg tactgagcca agataatgag ataacagaag gcgactttat tgttttccac 2160
tcaaaagcaa ttgaataagt tggcactcgt ttttaattga atgggaatga aataagctct 2220
aaaagtgttg ttaaaacgta atggcttttg tgttaattta aagaatttaa gtagttttga 2280
aagtatcatt attctttagg taatttttat tacattccaa atttaataaa tgactaattc 2340
gaaaaagtgt ttatttaatc aatgaatata tttcaagtaa gtttacttttt agtagcttgc 2400
caaatgtgag tttaaatatg tatgcataga actatatagt taaactgcta aactttacag 2460
ttaaactttc tgaacccacc aaaatggatg aacatcctcg tctgccgaag ggaactcgat 2520
gcacgtcatt ttgtttttca acaatccaga tccgtgcgct actccttggg cgagaaagta 2580
aacaaacgcc agctgatatg cgtcagaccc cccgggctca tcatcatctc accatttcag 2640
acatcccatg ccagcccgaa tcctcacgag aaactagacc agaccagggc gaactacata 2700
tgtggatgat gctaactgac actacggctg actcatgctg acagtgctca gacgctggat 2760
acagcccgca gacatccaac tcgtatccta tccgattctg ccccatatat ataaccctca 2820
gtcgatggct gggaggcaaa cagttgaggc cgtgccactt ggcagacaca tactacacac 2880
tccccggggg attcacgcat ccatacttaa acaccacttc atcactcatg ggcatcgaga 2940
tgaggtgtca ggacaggagg atcctgctac ctagcctact cctactaatc cttatgatcg 3000
gcggtgtcca ggccaccatg aagttgtgcg gccgcaaact gcccgaaact ctctccaagc 3060
tctgtgtgta tggcttcaac gcaatgacca agagaacttt gggtaggtgg gatttttctt 3120
gatataagga atactaaagt gccatatctc tttactttca cctaacacct gtagaccccg 3180
tgaacttcaa tcagatcgat ggcttcgaag accgttccct gctggaaaga ctgttgagtg 3240
atagttcggt tcagatgctc aagactcgac gtcttcggga tggagtcttc gacgagtgtt 3300
gcctgaagtc gtgcaccatg gatgaggtgc tgagatattg tgctgccaag ccgagaacgt 3360
aaacctcgta aacctattaa cccaatgacg acaactgcga tgattgaaat ggaatgaaag 3420

gacccgattg gggaaagcac tcacgtaatc atagttgtta agtcgttatc gaagcctact 3480
caattccaac tttggattta tgatatatat gcacatgtaa gagggatgta tgcgcataat 3540
ttatgatctg aaatcagaga caggcacgcg aaatgaatcg gaacacggga tgttatgcat 3600
ggtagatatg tatgattgtg cggggccaga atacatcgcc tgggtataaa ttattaaata 3660
aattatgtat tcaaactgct gcagattggc caacttgatt ggtaatgaaa cgggtattac 3720
attgattttt cattgtcgtt cattgcagtt aattatttat tgaacagcgg ccggatttct 3780
gtttgcaact atgttgaaaa ggaagctgtg atttttaac aaactctgtt cattgtaaag 3840
tttaaaatca ttccaattta atgccctcaa aacctacgct gaaatggtca gttttaaaac 3900
gatatttatt aatattttag ttaatttact aagattatcc gttttgcact tttaatgcct 3960
tgcatttggt aatgcgtgat tgttatttaa ggtctgcatg aatttagttg attccgttta 4020
ttttagcttt caaaatgtaa taatcttcta atttacaact acacagaacg attaaattat 4080
gagtattgct ataaaatcgg ccaaccgcga ctagaaatac tcgacttta aggtcaacat 4140
aaaagtaagt caatgttttg attataagat ttgatcaatt acttctttac ggatgatata 4200
atcatcgata aacgaagtac gaaaaaagct atgaactaaa atttggaaat ttcccacatg 4260
cgactaactt ttgaattgca attggattgc ctactgtatt aagacagaaa caagttttgg 4320
aaatgaatga atggtttaaa ttgtttcaag tttttttaag attttttttg ttttcaataa 4380
atttagtttt aatagaaaaa aagatatatt cattttagat ttctgaatac ttgtgttata 4440
tcgcttttta ttcaagtgta ataatcaaca tatatatcat ataatgataa taataaatgt 4500
aacgtcccaa attaataata atataaagta gcatttgcga ttgtttgcca aagcttaaag 4560
cagaatatat atttaatcca tttcgatcat tcgtaaagag taacatgcaa caagctgtaa 4620
aaaacatcga ttgtagtata tatgcacatg gttggtttgg aaccagatcc agagataatc 4680
gcgtcgacca ggtcagttgg ggtcaatgat tttcgcatta ggaggcctaa tttctgacca 4740
cggagcagta ctcccgcaga gccttcatat cacaggactt tttgcagcac ctctccacga 4800
ttccttgccg ttggcgagtt cgcctccgga cttcagccag ggaattgagt acacccccaa 4860
gatagctccc aggaaagagg gcacttcgca gcggttccga tatcgagtta tcctcctcct 4920
cgaactcctg gacaaactgc aggggattga gggcgtccag atcgctgtcg gcaccggcta 4980
ataaaaatcg tggatacaat gtagatctag caaagccagc ttgaggatct gcatccttgt 5040
aagaacttac gcatggcgcg cttgtgtgga atcacgggat tatactcctc gcacaccata 5100
ctcagcacct cgttgagctt ttcactgcag agcgttcctt gggccaactt cactgtggag 5160
ctggccagca aaatcacggc caccatcgag atgaaggaca aaggcttgct catggttatg 5220
ggtttactgc ttaggttgct ttacgatcaa atggattaag ttgggtcgag ccgggtcgaa 5280
agctaactga tgatgtttgg cccaaagtaa ctggcttata tactgcctcg taagaaactt 5340
aaactgggtc tgggtcgggg tcggtctctc ggggtcgggg tctggatcca cacacatgtt 5400
atcctcaaaa gtcaggttgt caaattgtgt taggatgcga tgagtgcatt ccggagttgg 5460
ctcttctctc taacgcctgg ctaaactcat tcaatgtcaa agctgactta tgcaaatggc 5520
tattggaaaa ttgtgggtgg tttttgggtg gctgtgtttg ggagaagaag ggctttgtgg 5580
gcgttttgct gtcagccaat taaacaattt atgtataaac agccaggccg tactaagccc 5640
tgcatttatg aataccaaat aagtccttgg tcttaaagtt acctcgcctt tacagcccgt 5700
ttgcctctac catttctacc ctatacttac caatccgcgc ctgggcgccc ggcaggccgg 5760
agtaggccaa caagaacccg agccagctga ttggagccag cagcatcctg gcaacgaatt 5820
acgcctcctt ggtacttttc ctttgactgt cttgtctttg ccgctcacac aaattcttct 5880
ttttgcactg tctactttta ttcattagtc aaagttggtg ctgcataaat aagtgattac 5940
gaattggatt acgaatgctg ttaggagaac gggtgtacat atagtatgta tgtgggaatg 6000
ccatgttcaa gtgttcgtat gtatgtatgt atgtgtatgc atgctgggta atgaatgtgt 6060
gtgtgttggc caagtgtcct atttcggtag acagtagatg gctaactcca agtagctgca 6120
ggtcttgacg cagcactcgt cgtagacgcc accggtcagg tgtctcctgt ggcgacgcat 6180
cttgatcagg acctcggatc cgtacagatt ggttagcagt ggactaaaag agtatcctgc 6240
cccgtccagt gtctgccaca tgctgctatc atcctgcacc tcctgctccg tgtcctcgtc 6300
gtcgtcgctg ttgcccagca agctttcacg tttccttggc agcgtattaa agccatgggg 6360
acacaccaca tccatggcat cggacagtgc ggggccgcag agtttgtggt ttccgggggg 6420
cagcaactgg tgaccactgc cagtcggcgt gaccattgcc attgcagcgg tgagcatggc 6480
tgcgatgagc agcgactgga gccgaaggcc atgtactgct gcaccgttgt gctggctaaa 6540
catcttggat atgcagtgaa tgctctgggc tgcaactggt atttatacca ctaaatcgga 6600
agcgctagct aatgcagttc aatggcctct tctgcagtct agcattgcag tggcatagca 6660
agccccacgg gcgtacaaac tgcaaatcct ttgatcaccc atgtttcagg taccgttttt 6720
cccctaaaaa tgcaaactct atttctagct ctactcccca atttggatgg aaaagcgatg 6780
cactgttgtt ttggtagttg gggtattgta ttgtatttct tagcaaatat cagttgtatc 6840
attacctata tctatctata ccaatagttt ggaatgtatt tgtaagacat ttttaagata 6900

ttcagaagag ttagccttat gggacttgct ctaaagtgtg aattgatgca cacagcttta 6960
tcgagcatag ttttcagtgt aatcaccgcc aaaaaatccg cccacttcaa agcataaccc 7020
gttcgcccaa cctgttacat tgccgctaag aggctctgac tgctgtcgat tgcgattacg 7080
attacgacca gatatctgtg gggcatgggg ataaggggta tgtgggccga tggctgacag 7140
tgtggcagcc tcattagcat gtcgtggcca ggaggaaagt atgcttcgat gaagctcctc 7200
cggcggcagt gtgcgaaatc gcttcgatca ccatcatcgc catcgccatg gccactcgat 7260
tgtcgagttg cacgcacggc gatgccaaca gttggttgcc agcgctgcac tcgaaacact 7320
cgcttcttcc caccgaccgc aaagtgccgg aaaagctaga aaaaaagcaa aaaaaaaagt 7380
ggaagaaaat tcgcgataga aaacggaaaa atcgaaacga acaaaaaaag tcggaataaa 7440
tcaaggaaac atggtgctcg acattaagat gtgccgattt gataatgtgc cctggggctt 7500
tcgcctggtg ggcggggcgg actacgatta tccgctgacg gtggttaagg ttaggcccga 7560
ttcgaaaaaa gaacgaaatc tatatgctgc aacccccacc cccccacgca tcacctcagc 7620
ccattcacct ggcggatgtt catagaccag tggaaaatat tgctcactat gcagctgatg 7680
aatcacattg gattaattcg atacgatacg ttcgaatcag ttttatttgt tcgattgcaa 7740
tattacgtaa cgccgcgatg cgtgtgtgtc cattcggatt tgctgcattg gcaaattagt 7800
taattaaagt aattcctctc gcttttgttt atctaatcga cagggccata catttcccgc 7860
taatgagccg cataatggca gcggcaataa acttattcaa attttaattg tgtttcgctg 7920
gcagttggtc ctttgtttgt gcataaattg catttggcaa ttcgcatttt gtaacattgt 7980
gttgacaatt cgcaaccagc aacaataaca aaaatacaat acatacaata ctatagcatc 8040
gtcgtaaatc cgaaacaaat gcgattttta attggcaaac tgctaagcgc ataaaacaaa 8100
tgaccgaaat gcgaggggcg ctaaaaaatc ccatcccttc gatacgaata aatcaattta 8160
agccgcagag tcaaggaagg gaggtcataa attgtttttg actttttggt tatttttttt 8220
ttaccgtttt acataaacaa attatgctat gggttatttt aaattccgat caatttataa 8280
aatgtttgtg ctttgggata tgcataccat gaaaaaatgg aagtttattg taaatgaatt 8340
attaacttca caagctggct gatagagaaa aaactgaaaa atgtccggaa tgttcttcat 8400
tccaatgaac tccctaaatt aacttagcta atttattcct tatactaata ctccgctttt 8460
aagaattcct tactacatgt tagagactca aaaagcacat ccttcgactc gagtccatat 8520
tactttatgg aatgtgccaa cacaccttca catattggct ctgcaaacac taaacaatcc 8580
ttggtaatct tttgaaaaac ctctgtttac actaccactc ttcgtcatgc tgctcgccac 8640
atacagtctg gtacatagat gtatggccca gctaagccca aagcctttgt tctataaata 8700
ttcgcaacct ccgacgatgt cgagtgcttt ttgctctgcg aattcaccgc tggaaattga 8760
ctctaccata agtgaaatgc aagagacccc tgggactgaa aggaaagacc ctcaacttgg 8820
ttgggtgaaa tggtggagtc tccaaccctc cacctgctcc ttgtgccaac cacttttttt 8880
tttttgcagt atttgcatta ctaagtcctc ttggcagtcg gtgtcgtgac tttctggtta 8940
tgaaccctgc tttctcataa cggaaacgaa aacaatcgcg tttatttgcc cacgaaagtg 9000
ttacaaaact gcctgaatta tgcaatagaa ttctttgaac agagtgctaa gatatttcgc 9060
attttgaagg cgaaacataa ttcatccata actattagtt tgatgaattc tcacttcgta 9120
tgctggcaat tttgaaggcc gaagtggcaa aaccatttta aatgaattcc tacaatttat 9180
atgctcaatt tgcgccaatt tgtctgtact ttattaccca caaaagccat aaagcttata 9240
ttgtatttag ttgttattgt tctcgcaaca ttttcacttt gatttatagt tgcgaaaata 9300
aatgttggca aatgcaaatt tttgaaacga tattctctta cacggtcatt tggtaccatt 9360
tccggaatat acatttacca attttccaaa aaaagagcca taaattgtat tatccaatta 9420
aaacgattag tccagtgggt ttttatttct aaaaatttaa ttttgtaatt agagaaaact 9480
attgtgaact ttggatgacc tagaaagttt cattagttgt acattatttt ttacccccgc 9540
tttaaaattg cagaacttct gtaaaaaaaa gcttttacaa gctatttaaa tattagtggt 9600
agagttgttt gaacatttat ctttttggaa tcagcaatat ttggttctct atccttacaa 9660
attaatcttc tctttaagtg aactcataac tttgtatata tgtacgtatg tagaatcatc 9720
cttcaactca aacatttcca tccggcgaaa caatgcaata tttgagtggt tagacatggg 9780
taattccact tgactcgctt aactgaagtg tcgttaatga gctgccactt ctactcgagc 9840
acccctcgtt ctgagccccc agcccccaca gatcctctgt agccccccca tctccttggg 9900
cattgtcttc ggttctgttg ggtttgtgcg tgattgtatg attgtctgtt cgggggtctg 9960
ggtctgctgg tctttgtttt gttgacattt ggcgcgcgtt ttatttttat gcacttagca 10020
cgcgacgtcg ccgttgtcgg ctaatagaaa tttccccatt atcgcatcgc atcccattgt 10080
attgtatctc ggctatctcg actatctcgg cctcgcactc attctatcgc cacattccca 10140
tatccgcatc tgaatcgggc gatatctagg cattcccatc tagatctaaa catgtccata 10200
tgcttcaacg ttaggtgacc gagggcagca ttgctgacga ggctggactg cgggtcgagg 10260
atatcatcgt gcgcatcaat gacacggctg ccacgcccct tacccacgac gaggcccacc 10320
gcctcattat gggcagtgga agcgtcttct attttggcgt ctaccggtga gctttcccat 10380

ccttctattc cataccattc cgtttcgttc tcttactgct cgtggtcgtg gtcctcgtcc 10440
ttgtccggtt cctttactac tactctttgt atcccatcca ccgaggacca tttatcacag 10500
ctgagcgggc taacaaccca agaacgtttc ctcatgcccc tgctcaaggt aatctacttg 10560
ttcaagaaat tcacaaacca caaacgacct cgagagaaat ggaaaaaata tgacaaattt 10620
tcgtgattta agaataaagt tctggaaaaa taaagcgctt tcttaaaaag ttgtctgggt 10680
aaaatgacat ttggttaata tatcataata gttaatttta ttataattat aaactaagaa 10740
aagttaaatt caaaaaccca cctagcccca ttagttttga aaattaccct accattttag 10800
aagcaattta tattatttga attaagtttg tatttcaact ttttcgggtt atgaataatt 10860
attcttagag tgtccccgaa accagggctc catctcaggt attccacgtt acggaattaa 10920
ttctaatgtt cataaattct gctcactttt tggtcacttg gatccatgtg cagggagaac 10980
gaggaggacg cttacgagtg cctaaagaag tttcccacga gcgagggttc gttgaccaag 11040
tcaccaatgc cgaccatttc accgtcgccg actccatcgc tgtcccagct gacggaaacc 11100
acaaatgccc gtactccgga 11120

<210> 8
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 8
ctaggaattc gatcgagcag gatgag 26

<210> 9
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 9
cacttctaga tcatcaggcg cagtag 26

<210> 10
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 10
cttcatcact catgggcatc gag 23

<210> 11
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 11
tgggttaata ggtttacgag gtt 23

<210> 12
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 12
gcttccgatt tagtggtata aa 22

<210> 13
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 13
ttcgtatgta tgtatgtatg tg 22

<210> 14
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 14
taaacccata accatgagca agc 23

<210> 15
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 15
tcagttgggg tcaatgattt tcg 23

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/28315

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) :C07H 21/04; C07K 14/00; C12N 15/85
US CL :435/325; 530/350; 536/23.1; 536/23.5
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/325; 530/350; 536/23.1; 536/23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	DURET et al. New insulin-like proteins with atypical disulfide bond pattern characterized in *Caenorhabditis elegans* by comparative sequence analysis and homology modeling. Genome Research. April 1998, Vol. 8, No. 4, pages 348-353, entire document.	1, 2, and 4-8
A	KONDO et al. Multiple gene copies for Bombyxin, an insulin-related peptide of the silkmoth *Bombyx mori*: structural signs for gene rearrangement and duplication responsible for generation of multiple molecular forms of Bombyxin. J. Mol. Biol. 1996, Vol. 259, pages 926-937, entire document.	1, 2, and 4-8

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 31 JANUARY 2000	Date of mailing of the international search report 22 FEB 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer ANNE-MARIE BAKER, PH.D. Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/28315

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LAGUEUX et al. cDNAs from neurosecretory cells of brains of *Locusta migratoria* (Insecta, Orthoptera) encoding a novel member of the superfamily of insulins. Eur. J. Biochem. 1990, Vol. 187, pages 249-254, entire document.	1, 2, and 4-8
A	SMIT et al. Growth-controlling molluscan neurons produce the precursor of an insulin-related peptide. Nature. 11 February 1988, Vol. 331, pages 535-538, entire document.	1, 2, and 4-8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/28315

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

Claim 1 was not searched on the basis of sequence because SEQ ID NOs were not provided for the sequences of Figure 8. Claim 1 was searched to the extent possible using keyword searching.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
1, 2 and 4-8

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

WEST
Dialog (file: medicine)
search terms: insulin-like, insulin-related, protein, family, Drosophila, C. elegans, gene, mutate, mutation, mutant.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1 and 2, drawn to a purified protein comprising an amino acid sequence of an A peptide domain and B peptide domain of a Drosophila insulin-like protein.

Group II, claim(s) 3, drawn to an antibody specific for a Drosophila insulin-like protein and not an insulin-like protein of another species.

Group III, claim(s) 4-8, drawn to an isolated nucleic acid encoding a Drosophila insulin-like protein, a host cell containing a vector encoding a Drosophila insulin-like protein, and a method of studying mutations in a Drosophila insulin-like protein gene.

Group IV, claim(s) 9, drawn to a knockout Drosophila melanogaster having a deleted D. melanogaster insulin-like gene.

Group V, claim(s) 10, drawn to transgenic animal containing a D. melanogaster insulin-like transgene.

Group VI, claim(s) 11, drawn to a compound screening assay for identifying a molecule that alters the expression level of a D. melanogaster insulin-like gene.

Group VII, claim(s) 12, drawn to a cell culture medium comprising a protein comprising at least 10 contiguous amino acids from a Drosophila insulin-like gene.

The inventions listed as Groups I-VII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The invention of Group I is distinct from the inventions of Groups II-VII because the inventions are drawn to materially different compositions and distinct methods. The protein of the invention of Group I is chemically, biologically, and functionally distinct from the antibody of the invention of Group II, the DNA and host cell of the invention of Group III, the knockout fly of the invention of Group IV, the transgenic animal of the invention of Group V, and the cell culture medium of the invention of Group VII. The invention of Group I is distinct from the invention of Group VI because the protein of the invention of Group I is not required to practice the method of the invention of Group VI, a compound screening assay that uses a transgenic fly cell.

The invention of Group II is distinct from the inventions of Groups III-VII because the inventions are drawn to materially different compositions and distinct methods. The antibody of the invention of Group II is chemically, biologically, and functionally distinct from the DNA and host cell of the invention of Group III, the knockout fly of the invention of Group IV, the insulin-like protein transgenic animal of the invention of Group V, and the cell culture medium of the invention of Group VII. The invention of Group II is distinct from the invention of Group VI because the antibody of the invention of Group II is not required to practice the method of the invention of Group VI, a compound screening assay that uses a transgenic fly cell.

The invention of Group III is distinct from the inventions of Groups IV-VII because the inventions are drawn to materially different compositions and distinct methods. The nucleic acid and host cell of the invention of Group III is chemically, biologically, and functionally distinct from the knockout fly of the invention of Group IV, the transgenic animal of the invention of Group V, and the cell culture medium of the invention of Group VII. The invention of Group III is distinct from the invention of Group VI because the nucleic acid and host cell of the invention of Group III is not required to practice the method of the invention of Group VI, a compound screening assay that uses a transgenic fly cell having a defined DNA construct encoding a reporter, not an insulin-like protein. The host cell of the invention of Group III contains a vector encoding an insulin-like protein, not a reporter molecule.

The invention of Group IV is distinct from the inventions of Groups V-VII because the inventions are drawn to materially different compositions and distinct methods. The knockout fly of the invention of Group IV is chemically, biologically, and functionally distinct from the transgenic animals of the invention of Group V and the cell culture medium of the invention of Group VII. The invention of Group IV is distinct from the invention of Group VI because the knockout fly of the invention of Group IV is not required to practice the invention of Group VI, a compound screening assay that uses a transgenic fly cell.

The invention of Group V is distinct from the invention of Group VI because the transgenic animal of the invention of Group V is not required to practice the method of the invention of Group VI, a compound screening assay that uses a transgenic fly cell. The invention of Group V is distinct from the invention of Group VII because the insulin-like protein transgenic animal of the invention of Group V is chemically, biologically, and functionally distinct from the cell culture medium of the invention of Group VII.

The invention of Group VI is distinct from the invention of Group VII because the cell culture medium of the invention of VII is not required to practice the method of the invention of Group VI, a compound screening assay that uses a transgenic fly cell.

Accordingly, Groups I-VII are not so linked by the same or a corresponding special technical feature as to form a single general inventive concept.